IN- SILICO INHIBITION OF GALECTIN-1 DURING HIV-1 PATHOGENSIS: A PHARMACOPHORE BASED VIRTUAL SCREENING, MOLECULAR DOCKING & QSAR STUDIES

Thesis submitted to

National Institute of Technology, Rourkela

For the Award of the Degree

Of

Master of Technology

In

Biotechnology

Submitted by

ARUN KUMAR MAURYA

(212BM2013)

Under the Guidance of

Dr. NANDINI SARKAR

Department of Biotechnology and Medical Engineering

National Institute of Technology

Rourkela

May 2014

Dr. NANDINI SARKAR

Assistant Professor

Department Of Biotechnology & Medical Engineering

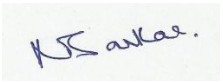
National Institute of Technology, Rourkela, Odisha, India

CERTIFICATE

This is to certify that thesis entitled **"*IN-SILICO* INHIBITION OF GALECTIN-1 DURING HIV-1 PATHOGENSIS: A PHARMACOPHORE BASED VIRTUAL SCREENING, MOLECULAR DOCKING & QSAR STUDIES" by ARUN KUMAR MAURYA (212BM2013)** submitted to the National Institute of Technology, Rourkela for the Degree of Master of Technology is a record of bonafide research work, carried out by him in the Department of Biotechnology and Medical Engineering under my supervision and guidance. To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/ Institute for the award of any Degree or Diploma.

DATE: 20/05/2014

PLACE: ROURKELA



Dr. NANDINI SARKAR

Assistant Professor

Department of Biotechnology and Medical Engineering

NIT Rourkela, 769008

ACKNOWLEDGEMENT

This project is by far the most significant accomplishment in my carrier and it would have been not possible without people who supported me and believed on me.

I consider it my privilege to express my gratitude, respect and sincere thanks to all those who guided and supported me in successful accomplishment of this project work. At first I would like to express my deepest sense of respect and indebtedness to my guide **Dr. NANDINI SARKAR**, Department of Biotechnology and Medical Engineering, NIT, Rourkela, for her consistent support, guidance, encouragement, precious pieces of advice in times, the freedom she provided and above all being nice to me round the year, starting from the first day of my project work up to the successful completion of the thesis work.

I would like to take the opportunity to express my gratefulness to all my respected Professors of my department specially **Dr. S. PAUL & Dr. A. THIRUGNANAM** for all the moments taken care by them throughout the program of study.

I extend my sincere thanks and respect to **Mr ARUN EVR**, who helped and motivated me during my project, without whose extraordinary support this work could not be completed.

I am also thankful to **Mr KARTHIC K, Ms SRUTHI UNNI**, my lab members, colleagues, juniors, and research scholars at NIT, Rourkela, who have helped me during my dissertation work and have been involved directly or indirectly in my endeavour

Last, but not the least, I would thank the Almighty, my parents, my brothers & sister whose dedicated and untiring effort towards me has brought me at this stage of my life.

Arun Kumar Maurya

M.Tech Biotechnology

Department of Biotechnology and Medical Engineering

NIT Rourkela, Odisha

ABSTRACT

Galectin-1 helps in HIV-1 pathogenesis by interaction between viral capsid protein gp120 and CD4 protein of susceptible helper T cell. Therefore, strategy to identify inhibitor for Galectin-1 will be of much importance in medical research. Here, structure based pharmacophore modelling, pharmacophore based screening, molecular docking, protein–ligand interaction fingerprints, binding energy calculations, and Quantitative Structure-Activity Relationship (QSAR) predictions were employed in a virtual screening strategy to identify novel inhibitors. A structure-based pharmacophore model was hypothesized that contained 6 pharmacophoric features as observed in crystal structure of galectin-1 and its inhibitor thiodigalactoside. The pharmacophore model was used to screen NCI, open database compounds release4 with 2, 46,374 active molecules. A total of 2233 hits were obtained from NCI database. The hits retrieved were further screened by molecular docking, protein–ligand interaction fingerprints, binding energy calculations. After that we predicted IC50 values of screened molecules. Based on these results, 6 molecules were predicted as potential inhibitors of galectin-1.

Key words: Galectin-1, HIV-1, Structure based phramacophore modelling, Molecular docking, QSAR, NCI database, IC50

LIST OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Introduction:

Human deficiency Virus (HIV)-1 is responsible for Acquired Immunodeficiency Syndrome (AIDS), and more than 25 million causality has been reported [1]. Cell mediated endocytosis helps HIV-1 to enter body cell. Susceptible helper T lymphocytes (T_H), target for HIV-1 entry express high level of CD4 protein on cell surface. The perquisite of HIV-1 is the infection of T_H cells by binding on to CD4 protein and so, it is called Lymphotrophic. Other kind of cells that binds HIV-1 virus includes macrophages, monocytes, dendritic cells, Langerhans cells, haematopoietic stem cells, certain rectal lining cells and microglial cells etc. These cells weakly bind to the virus particle than T_H lymphocyte because of low expression of CD4 proteins on the latter cell surface [2-4].

Adherence of virus to target CD4 cell occurs via specific interactions between the viral envelope glycoprotein (gp120) and the amino terminal immunoglobulin domain of CD4 protein. These interactions are enough for binding but are not sufficient for infection. So they expect additional cell surface proteins for initiating further fusion process of viral and target cell membrane [5]. Unlike other enveloped viruses, HIV-1 carries a limited number of envelop spikes, which are required for its adsorption to target cells [6, 7]. HIV-1 is supposed to circumvent this limiting element by exploiting the host's membrane adhesion molecules or soluble proteins that can promote attachment of viral particles to target cells [8-15, 17, 18]. One of the host molecules exploited by HIV-1 is galectin-1, that has been reported to enhance both HIV-1 binding and infectivity in CD4+T cells and macrophages by increasing viral adsorption to target susceptible cells [14-16]. Since galectin-1 is profusely discovered in organs rich in CD4+ T cells, such as lymphoid tissues and tissues encompassing the lamina propria of the genital and gut mucosa [19-21], it may act as a significant role in HIV-1 transmission. Thus, specific inhibition of galectin-1 could exemplify an interesting approach to chemically interfere

with HIV-1 propagation and to maximize the efficacy of HIV-1 attachment/entry inhibitors.

Galectins are soluble glycan-binding proteins holding one or two carbohydrate recognition domains (CRDs), specified by conserved peptide sequences of approximately 130 amino acids that are responsible for their β-galactoside-binding [22]. In spite of the similarity of their CRDs, each galectin exhibits a unique ligand preference that depends on the β-galactoside structure as well as its substitutions [23]. For example, galectin-3, as opposed to galectin-1, does not bind very efficiently to HIV-1 or its primary cellular receptor, CD4 [16, 17]. This indicates a specific interaction between galectin-1 and HIV-1, which could be relevant for AIDS pathogenesis. So far, 15 galectins have been identified in mammals. Classification of galectins relies on the structural presentation of their CRD (i.e., prototype, chimera, and tandem repeated [24, 25]. For example, galectin-1 is a prototype galectin, while galectin-4 is a member of the tandem-repeat type and galectin-3 is the sole representative of the chimera type.

In the context of HIV-1, it has been reported that galectin-1 is competent to cross-link molecules found on the exterior of both virions and target cells, thus consequence in a significant enhancement of HIV-1 infection [14-17]. Due to the characteristic ability of galectin-1 to specifically bind to clustered complex type glycans on HIV-1 and increase virus infectivity [16]. New inhibitors that interfere with galectin-1-mediated interactions could be clinically applicable and the look for a specific inhibitor for galectin-1 carry on.

Therefore, we achieved to find synthetic compounds derived from the lactoside or galactoside molecule that could peculiarly inhibit galectin-1 activity in a cellular model of HIV-1 infection by altering their attached aglycone structures of lactoside or galactoside. While lactose has been used to inhibit galectin's activities in many probes, it is known to target every β-galactoside binding lectin and requires high concentrations (at

least 10 mM) to be effective. Variations in aglycone structures, which modify the charge density or multivalency of lactoside derivatives, allow lactosides to have more stable and specific interactions with the CRDs of selected galectins [57].

1.2 OBJECTIVES:

To find potential inhibitors of Galectin-1 using structure-based virtual screening methods and also predict inhibitory concentration (IC50) of leads.

1. To develop a small molecules database
2. To generate structure based pharmacophore model
3. To screen compound database with pharmacophore hypothesis
4. To prepare protein target using GROMACS
5. To perform molecular docking analysis with hits
6. To generate and validate QSAR model
7. To predict IC50 value of potential inhibitors

1.3 PLAN of WORK

SL No.	Works	3rd Semester		4th semester	
		Mid Semester	End Semester	Mid Semester	End Semester
1	Literature review & survey				
2	Prepare/select target protein molecule				
3	Binding site prediction				
4	Small molecules database generation				
5	Prepare structure based pharmacophore model				
6	Pharmacophore based virtual screening				
7	Virtual high throughput screening e.g. Molecular docking analysis				
8	QSAR Studies				
9	Documentation				

CHAPTER 2

LITERATURE REVIEW

2.1 Human Immunodeficiency Virus (HIV):

The acquired immune deficiency syndrome (AIDS) was first reported in patients in the USA in 1981. Later characterization of the principal causative agent, human immunodeficiency virus type 1 (HIV-1), uncovered that it was a retrovirus. As strains of HIV-1 were sampled from around the world, it became manifest that they exhibit extremely high genetic heterogeneity and that analysis of the evolution of this diversity can reveal insights into the prehistory of the virus [26].

During HIV-1 was first described, the closest known relative was visna, a virus from sheep that is the prototypic member of the genus Lentivirus. After that lentiviruses were soon found in other primates, and a second virus (HIV-2) was found tainting humans. The viruses from non-human primates were named simian immunodeficiency viruses (SIVs). Among the first species incurred to be naturally infected were African green monkeys (Chlorocebus species), where the preponderance of infection is much more (greater than 50% of adults) and natural infections appear to be non-pathogenic. The number of different SIVs distinguished has increased steadily over the past 20 years. Presently, around 40 different primate species have been discovered to harbour SIVs, though information regarding prevalence and pathogenicity is lacking for most. So far, SIVs have only been found naturally infecting primates in sub-Saharan Africa, though the extent to which Asian or new world primates have been reviewed is unclear. Where multiple strains of SIVs have been characterized from a single species, they mostly form a monophyletic clade, suggesting that the great majority of transmissions are intraspecific. The primate viruses as a entire, including HIV-1 and HIV-2, form a distinct clade within the lentiviruses, showing that humans acquired their infections from other primates [27]. Phylogenetic analyses of these primate lentiviruses have provided outstandingly detailed insights into the evolutionary origins of the human viruses.

2.2 Life Cycle of HIV 1:

The HIV-1 transmission action [28] is inducted by the binding of the viral envelope glycoprotein complex (Env; gp120, and gp41) to CD4, which is found on the surface of susceptible cells (Fig. 1-1). The transmembrane constituent of the envelope protein, gp41, can then initiate fusion of the viral membrane with the plasma membrane of the host cell. After fusion and transfer of the viral capsid in the cytoplasm, decapsidation directs to the release of viral enzymes, proteins, and genomic RNA inside the cell (Fig. 1-2, Fusion). Reverse transcription of the viral genomic single-stranded positive RNA is then initiated to yield a double-stranded proviral DNA to be imported in the nucleus and integrated into host chromosome (Fig. 1-3 and 1-4, Reverse Transcription and Integration).

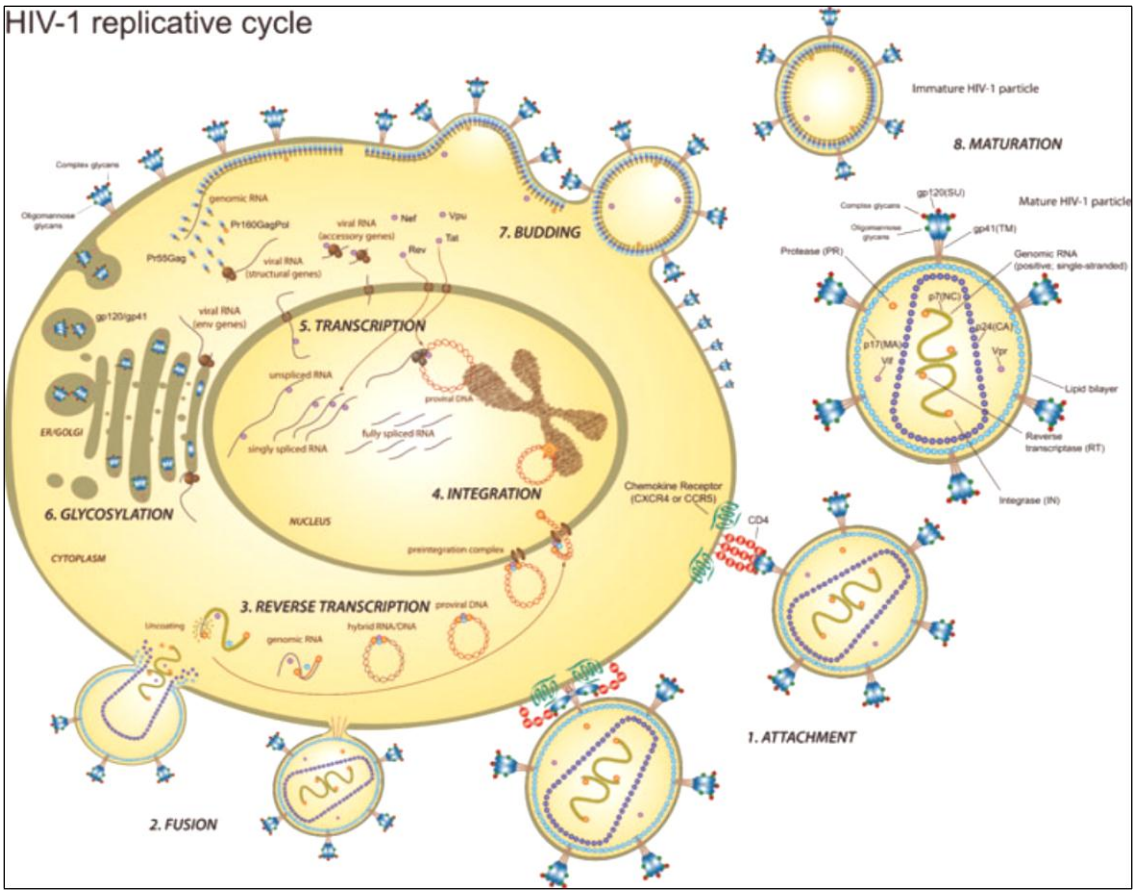


Figure 1. HIV-1 Replication cycle [57]

2.3 Role of Host Molecules in HIV 1 Infection:

HIV-1 is so recognized by dendritic cells (DCs) and Langerhans cells (LCs) via particular C-type lectins, such as DC-specific ICAM3-grabbing non-integrin (DC-SIGN), DC immune-receptor (DCIR), and Langerin [9, 11, 46]. DC-SIGN is extremely expressed in monocyte-derived DCs and dermal and mucosal DCs but not in LCs [47]. DC-SIGN binds to oligomannose-type glycans of gp120 and helps both direct infection of DCs (cis-infection) as well as DC-mediated HIV-1 transfer toward CD4+ T cells (trans-infection) [9, 48]. DCIR is expressed in antigen presenting cells, including myeloid and plasmacytoid DCs, also macrophages, but very few in LCs [49]. This lectin has recently been suggested to be related in both cis- and trans-infection by HIV-1 using monocyte-derived DCs [11].

Instead of C-type lectins, two classes of host proteins, integrins and syndecans, are known to help HIV-1 infection. Two integrin family members, α4β7 and αLβ2 are listed as leading HIV-1 entry into susceptible cells. CD4+ T cells express αLβ2 (LFA-1), which promotes HIV-1 attachment by binding to ICAM-1, an adhesion molecule adopted by HIV-1 during the budding process. [50-52]. One more family that promotes HIV-1 pathogenesis is syndecans [51]. Syndecans belong to a class of proteoglycans that contain covalently linked, heparan sulfate glycosaminoglycans, to which gp120 directly attaches [53, 54]. In fact, HIV-1 retains its infectivity for as long as a week when bound to cell-surface syndecans and can be readily transmitted upon contact with CD4+ T cells [46, 53, 54]. In addition to these host membrane bound proteins, the host soluble lectin galectin-1 also imparts to HIV-1 binding to CD4+ susceptible cells and promotes HIV-1 infection [14-17, 56] (Figure 2).

The possible role played by galectin-1 in the pathogenesis of HIV-1 infection is described later.

2.4 Galectins:

Galectins are a phylogenetically conserved family of lectins explained in 1994 as a shared consensus of amino-acidsequences of about 130 amino acids and the carbohydrate recognition domain (CRD) responsible for β-galactoside binding [22]. Fifteen mammalian galectins have been recognized to date. While some of these galectins have one CRD and are biologically active as monomers (galectins-5, -7, -10), as homodimers (galectins-1, -2, -11, 13–14, -15) or as oligomers that aggregate though their non-lectin domain (galectin-3); others contain two CRDs connected by a short linker peptide (galectins-4, -6, -8, -9, -12). While the CRDs of all the galectins contribute an affinity for the minimum saccharide ligand N-acetyllactosamine- a common disaccharide found on many cellular glycoproteins-individual galectins can also recognize different modifications to this minimum saccharide ligand and so examine the fine specificity of certain galectins for tissue- or developmentally-specific ligands [29]. Location studies of galectins have established that these proteins can isolate into multiple cell compartments in function of the status of the cells in question [30, 31]. Although galectins as a whole do not have the signal sequence required for protein secretion through the usual secretory pathway, some galectins are secreted and are found in the extracellular zone [32].

2.5 Galectin-1:

Galectin-1, a ubiquitous member of the galectin family, binds favourably N-acetyllactosamine but can also obligate a variety of β-galactosamides with modifications in the non-reducing end, the glycosidic linkage, or even longer oligosaccharides [33-36]. Galectin-1 acts as a noncovalent homodimer [34], in which each monomer is formed by 11 antiparallel β-strands generating a β-sandwich, with one CRD per monomer. The

galectin-1 CRDs carry the following residues: His-44, Asn-46, Arg-48, His-52, Asn-61, Trp-68, Glu-71, and Arg-73.

Galectin-1 is presented in the thymus and by lymphoid parenchymal epithelial cells, endothelial cells, trophoblasts, activated T and B cells, macrophages, follicular DCs, and CD4+ CD25+ regulatory T cells.[20, 37-41, 21]. Within subtypes of CD4+ T cells, galectin-1 is highly expressed and secreted by T_H1 cells [42]. Significant expression of galectin-1 is found in the lamina muscularis mucosae, just beneath the epithelium and the lamina propria, where HIV-1 prone CD4+ T cells can be found [19, 43, 44] galectin-1 expression is intense in HIV-1-infected CD4+ T cells compared to non-infected by stander cells or untreated cells. [45]

Recently experimented that galectin- 1, but not galectin-3, strongly binds to recombinant CD4, the main host receptor for HIV-1. Further, galectin-1 increases HIV-1 binding to CD4- expressing cells [16]. Thus, the discriminating recognition of CD4 by galectin-1 but not galectin-3 is likely due to the difference in their binding preferences for bi-antennary complex-type glycans presented on the surface of CD4.

Thus, while further studies are justified, galectin-1 can be listed as one potential host factor that could directly help to HIV-1 transmission.

2.6 Galectin-1 Inhibitors:

Galectin-1 directly binds to HIV-1 virus particles. These activities of galectin-1 in HIV-1 infection are inhibited by lactose, a β-galactoside-containing saccharide, but not mannose, confirming that galectin-1 distinguish β-galactoside residues expressed on HIV-1 [14-17]. HIV-1 entry strictly relies on the interaction between gp120 and CD4/coreceptor, pointing that galectin-1 only assists the initial virus binding event without affecting the rest of the HIV-1 entry process. Therefore, we endeavoured to find synthetic compounds derived from the lactoside or galactoside molecule that could specifically inhibit galectin-

1 activity in a cellular model of HIV-1 infection by altering their attached aglycone structures of lactoside or galactoside.

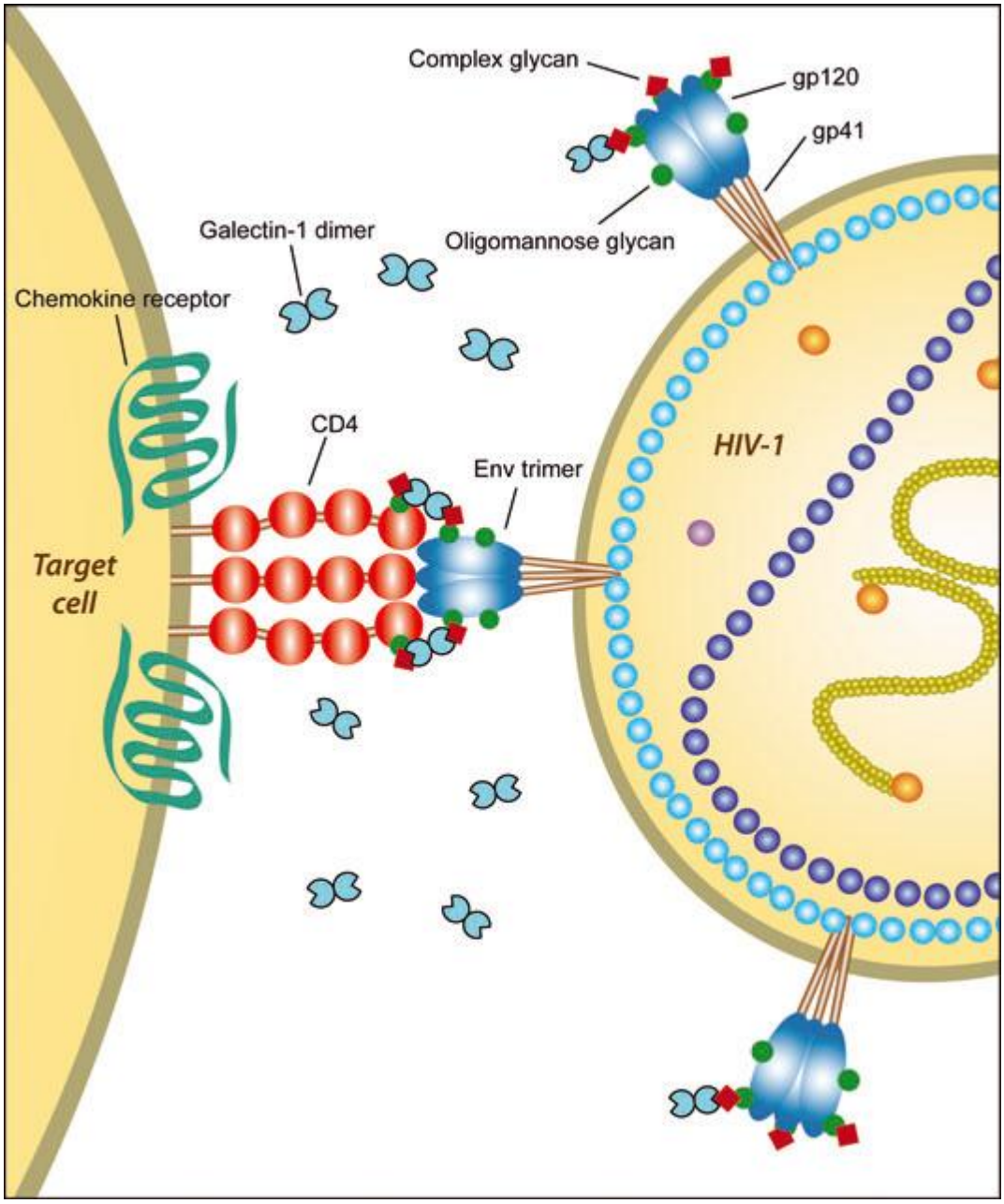


Figure 2. Galectin-1–mediated HIV-1 adhesion to CD4+ T cells [57]
[Sato *et al. Galectins and HIV-1 infection Ann. N.Y. Acad. Sci. 1253 (2012) 133–148]*

2.7 Computer Aided Drug Discovery (CADD):

Computational techniques in drug discovery and development process is rapidly acquiring in popularity, implementation and appreciation. Different terms are being applied to this area, including computer-aided drug design (CADD), computational drug design, computer-aided molecular design (CAMD), computer-aided molecular modeling (CAMM), rational drug design, in silico drug design, computer-aided rational drug design. Term Computer-Aided Drug Discovery and Development (CADDD) [58].

Fast expansion in this area has been made possible by advances in software and hardware computational power and sophistication, identification of molecular targets, and an increasing database of publicly available target protein structures. CADDD is being utilized to identify hits (active drug candidates), select leads (most likely candidates for further evaluation), and optimize leads i.e. transform biologically active compounds into suitable drugs by improving their physicochemical, pharmaceutical, ADMET/PK (pharmacokinetic) properties. Virtual screening is used to discover new drug candidates from different chemical scaffolds by searching commercial, public, or private 3-dimensional chemical structure databases. It is intended to reduce the size of chemical space and thereby allow focus on more promising candidates for lead discovery and optimization. The goal is to enrich set of molecules with desirable properties (active, drug-like, lead-like) and eliminate compounds with undesirable properties (inactive, reactive, toxic, poor ADMET/PK). In another words, in silico modelling is used to significantly minimize time and resource requirements of chemical synthesis and biological testing [59].

Applications and benefits of CADDD have been reviewed and demonstrated in growing number of publications and supported by examples of drugs derived from the in silico approach [59–65]. Virtual screening has been shown more efficient than commonly used empirical screening. Shoichet reported that ligand discovery i.e. hit rates (number of compounds binding to a target divided by number of compounds tested) is greater in virtual screening by 2 or 3 orders of magnitude than in empirical screening [66].

Pharmacophore library screening followed by docking represent complimentary screening methods (Figure 3) with the combination allowing optimum results [59]. Commonly, this screening approach is introduced by a prior filtering of virtual databases (e.g. physicochemical, ADMET/ PK, stability, reactivity, toxicity, drug-like properties, etc.) [59, 67–71]. This combination of screening methods has been successfully employed in designing new hits and leads.

Typically, this approach involves virtual screening (pharmacophore plus docking) of virtual chemical structure libraries containing hundreds of thousands of compounds.

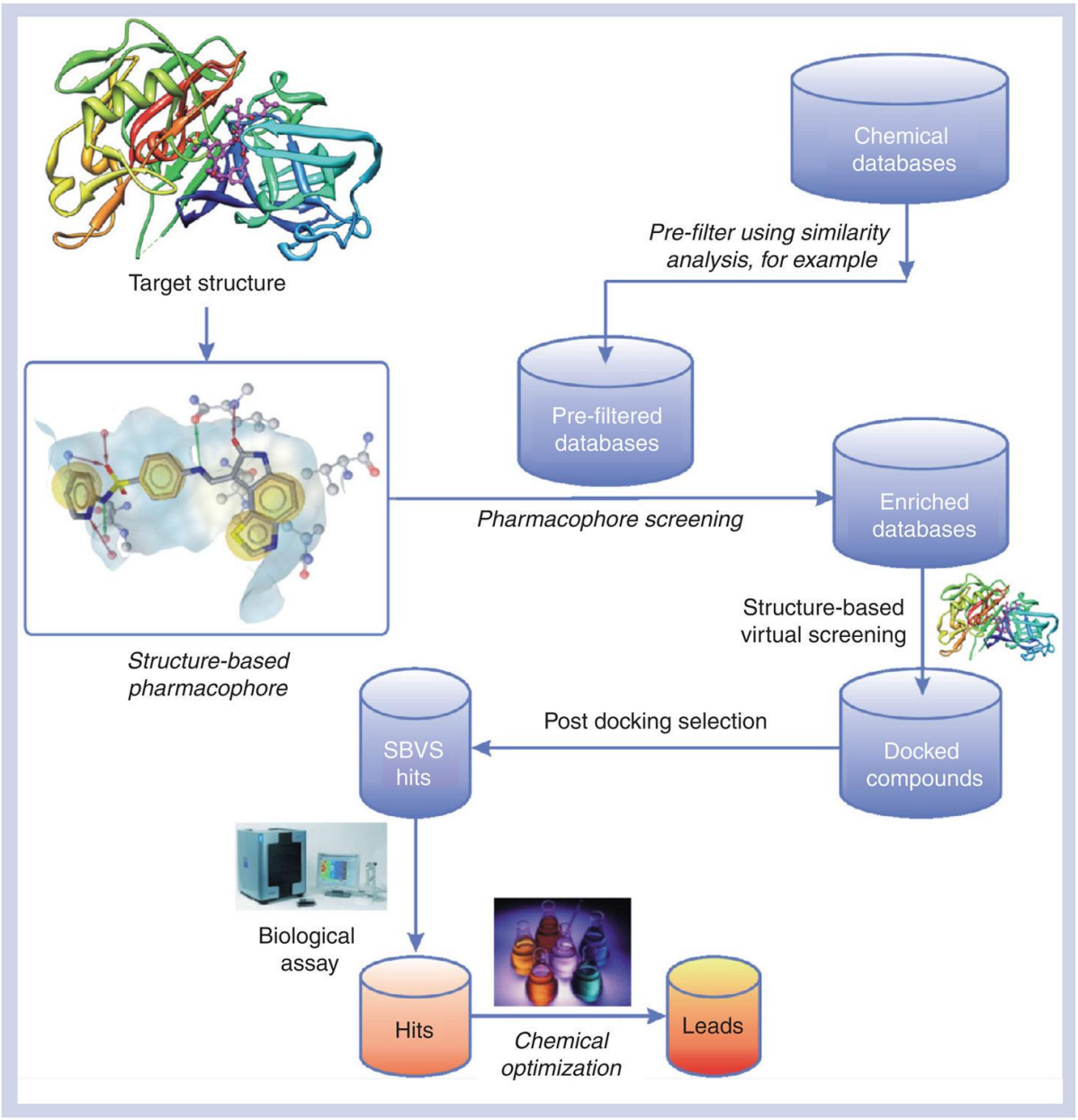


Figure 3. Pipeline of structure-based virtual screening in drug-discovery [58].

CHAPTER 3

MATERIALS & METHODS

3.1 BIOINFORMATICS TOOLS & SOFTWARES:

1. National Cancer Institute Open Database Compounds-Release 4 File Series - May 2012
2. Developmental Therapeutics Program NCI/NIH
3. LigandScout v3.1 software
4. Protein Data Bank (PDB)
5. Chimera v1.8 software
6. GROMACS software v4.6.5
7. National Center for Biotechnology Information (NCBI)
8. Autodock-Vina
9. MGL Tools v1.5.6
10. LigPlot+ v1.4.5 software
11. CHEMBL database
12. McQSAR software
13. PaDEL-Descriptor

3.2 PREPARATION OF SMALL MOLECULES LIBRARY:

Compound library database was downloaded from National Cancer Institute Open Database Compounds-Release 4 File Series - May 2012 [72] as provided by Developmental Therapeutics Program NCI/NIH (http://dtp.nci.nih.gov/). Open NCI database contained 265,242 structures in SDF format. This SDF format file was processed in LigandScout accessible format (.ldb format), using LigandScout v3.1 software [73]. LigandScout provides two type of confirmation analysis: Omega FAST and Omega BEST quality analysis. Here we were used Omega FAST to generate library database.

3.3 GENERATION OF STRUCTURE-BASED PHARMACOPHORE MODEL:

Pharmacophore modeling is one of the most frequently applied and valuable methods of discovering novel scaffolds for various targets, so this approach was used in this work to find novel inhibitors of Galectin 1. Two different types of pharmacophore procedures can be used to discover the most potent leads: (i) structure-based and (ii) ligand-based. Here we used structure-based pharmacophore generation, which depends protein-ligand/inhibitor complex structure. The protein complex (receptor) was selected from the Protein Data Bank [74] to generate pharmacophore model. Many Galectin 1 complexes have been reported; among Galectin-1 and thiodigalactoside complex (PDB ID: 3OYW, Figure 4) was selected as the receptor based on the resolution of the complex as well as the deposited date. Protein-ligand complex was submitted to LigandScout vs.3.1 software and it present protein and ligand interaction as well as excluded volume of 3D structure of protein. LigandScout automatically generated pharmacophore model which showed total 21 chemical features, e.g. hydrogen bond donor (HBD) and hydrogen bond acceptor (HBA). We defined phrarmacophore hypothesis that contain 6 feature (2HBD & 4HBA).

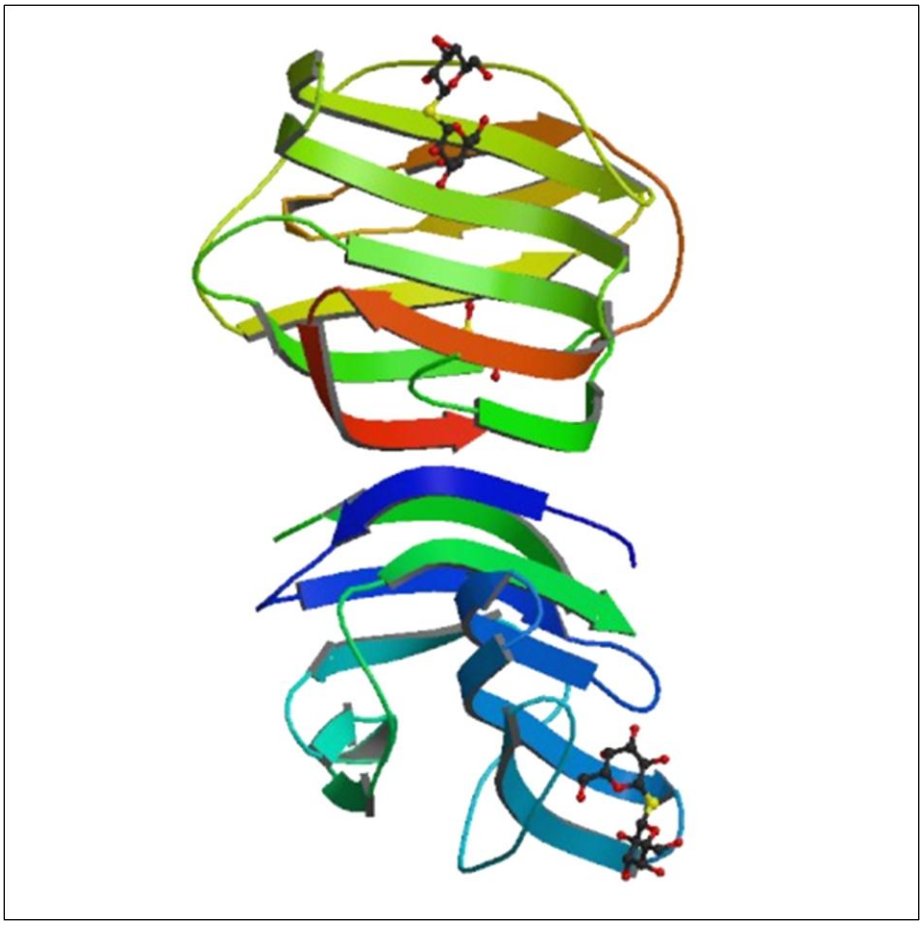


Figure 4. Crystal structure of human galectin-1 in complex with thiodigalactoside [74].

3.4 PHARMACOPHORE-BASED VIRTUAL SCREENING:

Selected pharmacophore hypothesis was used for virtual screening to retrieve a novel compound for Gal-1 inhibition from NCI database (2, 46,274 compounds). The screened compounds were filtered based on pharmacophore fit-score (≥60) and activity. The screened hits that satisfied the filter were selected for molecular docking studies to obtain the interacting leads with the active residues of Gal-1.

3.5 PREPARING THE RECEPTORS:

Galectin-1 (receptor) complex with inhibitors e.g. lactose (1W6O), galactose(1W6M), N-acetyllactosamine (1W6P), thiodigalactoside (3OYW) was downloaded from Protein Data Bank. We used 1W6O x-ray crystallographic structure with higher resolution $1.90A^0$ for molecular docking. Protein was visualized and ligands were removed using Chimera v1.8 software [75]. We also edited (renumber from 1) and minimized the structure in Chimera v1.8 software. We performed protein energy minimization and simulation using GROMACS [76] software v4.6.5.

3.5 BINDING SITE PREDICTION:

Gal-1 carbohydrate recognition domain's (CRD) residues was preferable target. Gal-1 CRD contain: His-44, Asn-46, Arg-48, His-52, Asn-61, Trp-68, Glu-71, and Arg-73 which was confirmed from literature and using National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). We kept Arg-48 as center residue in grid box and started molecular docking.

3.6 MOLECULAR DOCKING:

Molecular docking is one of the best methods used in the structure based drug design process. Hence, docking was carried out, using Autodock-Vina v1.2 [77] to find the most suitable orientation and interactions (hydrogen bonds and hydrophobic interactions) of each lead at the protein's active residues. The Protein-ligand complex were visualized in Chimera vs1.8

software [75]. Further, the grid was set with co-ordinates x=4.584, y=55.058, z=16.511 and volume 40*40*40. Maximum 20 number of mode was defined. The automated docking method was performed and docked molecules were filtered based on binding energy (> 7.0 Kcal/mol) and then analyzed.

3.7 DOCKING ANALYSIS:

Protein and ligands interaction were analyzed, using MGL Tools 1.5.6 (http://mgltools.scripps.edu/downloads/MGLTools-1.5.6/) Vina output file were loaded in analysis window of MGL tool software followed by macromolecule. Then we have seen hydrogen bond and hydrophobic interactions between macromolecule's residues and ligands for each 20 clusters. After that we plotted 2D image of best docked results using LigPlot+ [78] v 1.4.5 software.

3.8 QSAR STUDIES:

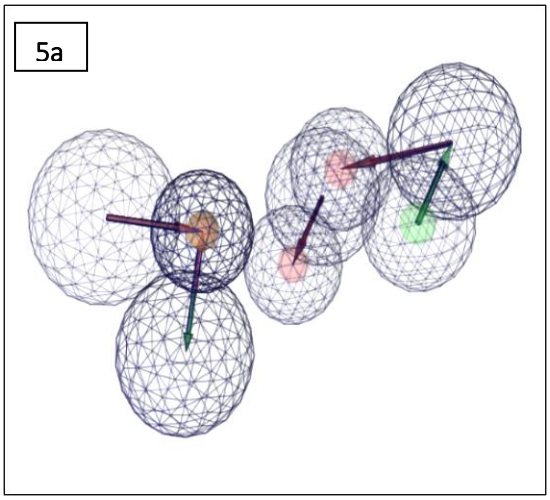
Quantitative Structure-Activity Relation (QSAR) studies was performed in McQSAR software [80]. For QSAR, we defined training and test set molecules. 35 compounds with experimental activity values e.g. IC50 were found and downloaded from CHEMBL database [79]. LigandScout software were used to cluster the 35 compounds and chose as training. Molecular descriptor for training set and test set compounds were generated using PaDEL-Descriptor [81].These training and test set descriptors were further used to build and correlate the QSAR model, using McQSAR software. Total 316 filtered dock molecules was used as query test set compounds. Then we predicted structure activity relationship (SAR) for test set compounds using McQSAR.

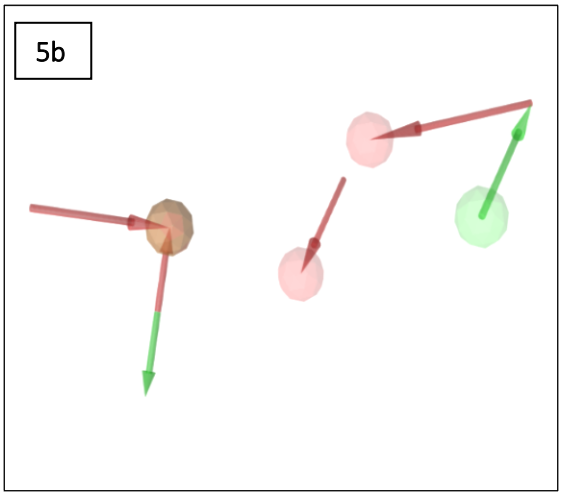
CHAPTER 4

RESULTS & DISCUSSION

4.1 GENERATION OF STRUCTURE BASED PHARMACOPHORE MODEL:

These days computer aided drug discovery is faster growing research field which escape hard work and save our time. CADD contain many tools and technologies, one of them is pharmacophore modelling. Pharmacophore model represent drug like chemical features (HBD, HBA, hydrophobic properties, aromatics group, polar charged group etc.). Gal-1 and thiodigalactoside complex x-ray crystallographic structure (3OYW) was submitted in LigandScout software and automatically structure based pharmacophore was generated. The generated pharmacophore having 21 chemical features (HBD & HBA) and exclusion volume. We removed exclusion volume and built a phamacophore hypothesis with 6 pharmacophoric features (2HBD +4HBA) [figure 5a, b, & c]. The pharmacophore hypothesis was validated by analysing the tolerance and Cartesian coordinates of pharmacophoric features (Table-1).





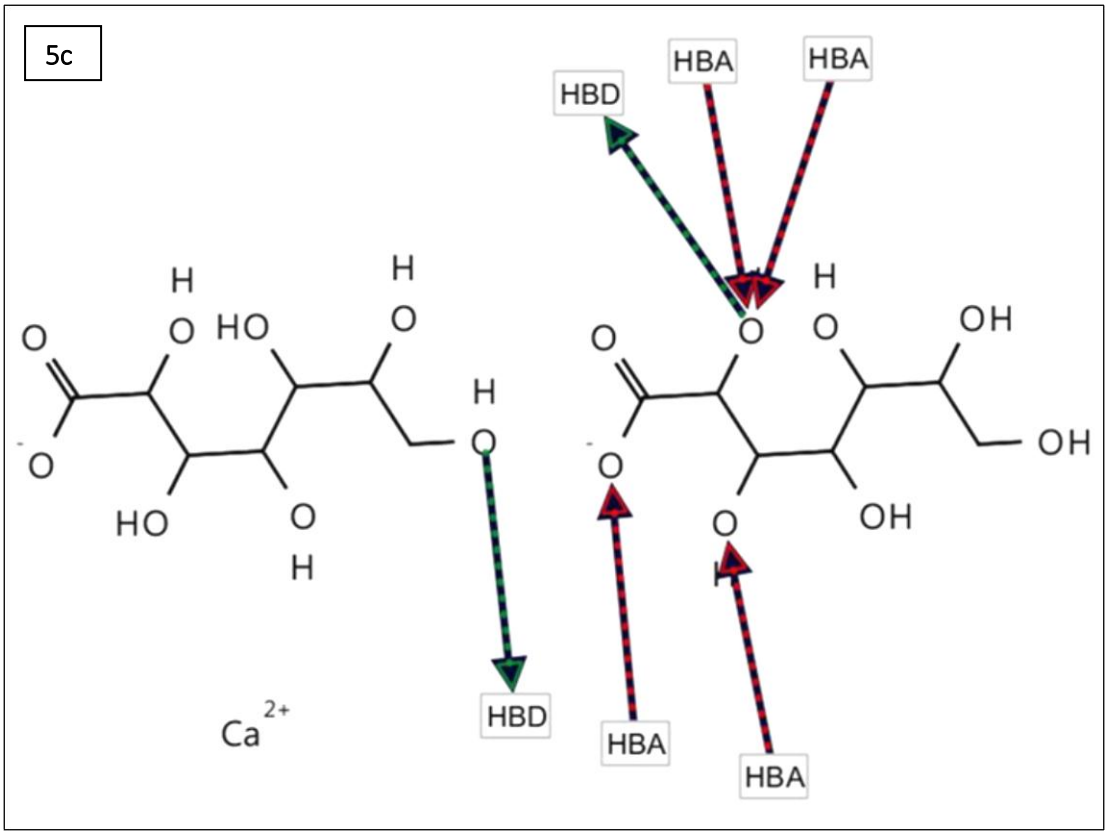


Figure 5a b & c. Pharmacophore Hypothesis with selected 6 chemical features (Red = HBA: Hydrogen bond acceptor & Green = HBD: Hydrogen bond donor)

Table 1. Tolerance & Cartesian Coordinates of Pharmacophoric Features

Sl.No.	Feature Type	Origin Position(x,y,z)	Target Position(x,y,z)	Tolerance	Weight
1	HBA	22.92,4.33,14.05	21.10,2.37,16.31	1.5	1.0
2	HBA	21.63,-0.07,14.75	21.97,0.13,17.74	1.5	1.0
3	HBA	21.27,-1.71,21.04	19.49,-0.52,19.06	1.5	1.0
4	HBA	17.49,-0.66,21.09	19.49,-0.52,19.06	1.5	1.0
5	HBD	19.49,-0.52,19.06	21.27,-1.71,21.04	1.5	1.0
6	HBD	23.14,4.10,17.34	22.92,4.33,14.05	1.5	1.0

4.2 PHARMACOPHORE BASED VIRTUAL SCREENING:

Many compound databases (e.g. DrugBank, PubChem,CHEMBL and NCI etc.) are freely available. We preferred and downloaded NCI database which contain anti-cancerous and anti-HIV compounds. We processed database in LigandScout accessible file format, using LigandScout v3.1 software. We applied Omega FAST setting (Figure 6), that generate maximum 25 confirmations of molecules. Then we screened NCI database with our pharmacophore hypothesis. Pharmacophore based virtual sceening of NCI database have given total 2233 hits out of 2, 46, 274 active compounds. These 2233 hits had good pharmacophore score (PS≥60) (Table-2). We used all 2233 hits for molecular docking studies.

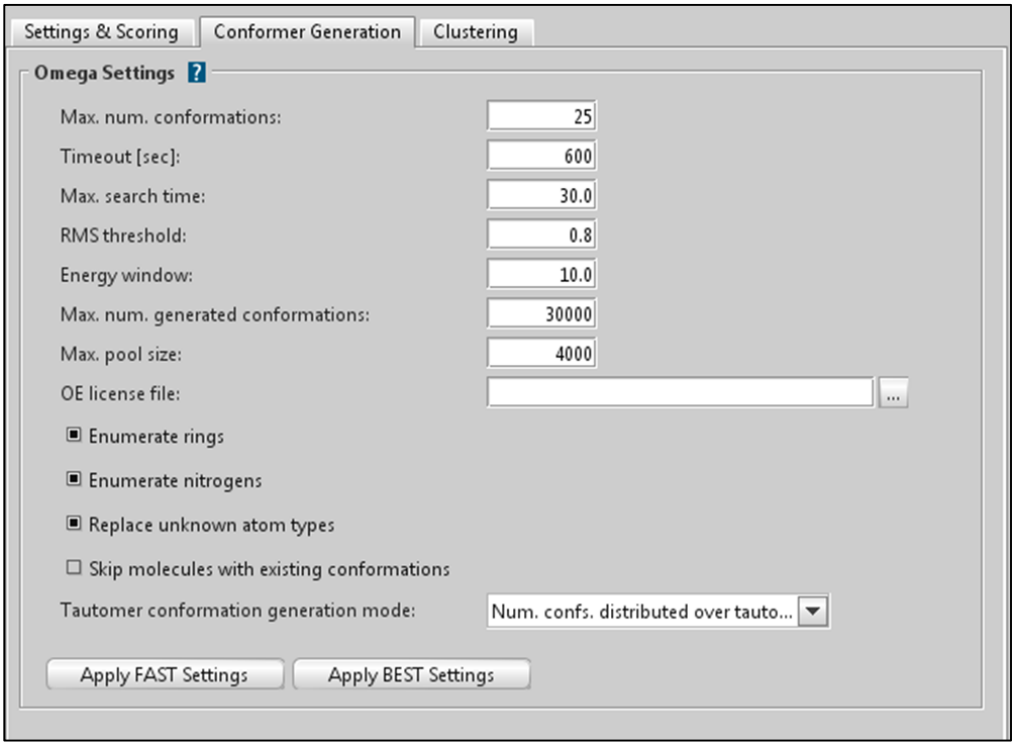


Figure 6. OMEGA FAST Setting

Table 2. NCI Database Hits with good pharmacophore score (PS≥60).

Index	NSC_Name	Pharmacophore-Fit Score
1	997	68.8
2	1017	63.86
3	996	68.79
4	1009	63.86
5	1219	67.36
6	1218	68.53
7	1220	66.33
8	1224	65.86
9	1368	67.4
10	1380	65.44
11	1382	66.67
12	1679	68.49
13	1667	66.76
14	1670	68.49
15	1665	66.98
16	1671	68.74
17	1684	68.04
18	1687	65.72
19	1688	67.09
20	1692	68.21
21	1683	66.43
22	1749	66.86
23	1751	66.86
24	1752	67.33
25	1948	68.46
26	1947	68.57
27	1952	68.81
28	1945	66.68
29	1960	67.13
30	1951	66.89
31	1961	68.78
32	1955	67.41
33	1959	65.63
34	1962	67.07
35	2027	66.71
36	2056	67.66
37	2342	64.9
38	2555	66.15
39	2564	64.09
40	2568	68.53
41	2567	66.19
42	2680	66.05
43	2833	66.92
44	3993	64.83
45	4010	
46	4036	66.79
47	4243	66.83
48	4245	66.38
49	4348	66.75
50	4320	68.79

Index	NSC_Name	Pharmacophore-Fit Score
51	4454	66.72
52	4722	66.25
53	5112	67.69
54	5159	65.3
55	5554	68.2
56	5548	66.96
57	5751	68.57
58	6196	65.92
59	6835	66.27
60	6836	66.77
61	7238	66.47
62	7391	65.7
63	7525	68.79
64	7531	66.3
65	7530	66.11
66	7533	68.78
67	8099	67.84
68	8512	65.18
69	8689	66.07
70	8878	65.36
71	9147	64.84
72	9161	67.52
73	9248	67.35
74	9250	66.5
75	9684	66.07
76	9892	66.09
77	10184	64.27
78	10373	65.99
79	10981	66.12
80	11075	66.33
81	11794	66.24
82	11796	64.65
83	12161	64.96
84	12519	66.1
85	14044	68.11
86	14161	65.45
87	15193	67.58
88	15606	64.85
89	15899	65.13
90	15780	66.18
91	15988	68.62
92	16409	64.98
93	16671	65.73
94	16653	66.3
95	16673	64.25
96	16757	66.92
97	16866	66.05
98	17250	66.41
99	18260	66.2
100	18695	67.1

Index	NSC_Name	Pharmacophore-Fit Score
101	18721	64.79
102	19434	67.43
103	19510	68.16
104	19772	65.48
105	19771	65.73
106	19777	67.08
107	20591	68.83
108	20599	68.83
109	20606	67.87
110	20615	66.8
111	20629	65.64
112	20634	67.64
113	20641	67.15
114	20664	66.68
115	20720	65.41
116	20721	65.43
117	21277	68.53
118	21402	66.78
119	21563	66.85
120	22079	65.17
121	22941	65.89
122	23190	65.51
123	23587	66.28
124	23789	67.24
125	23471	66.1
126	24187	65.82
127	24998	67.47
128	25270	67.37
129	25269	67.53
130	25271	67.4
131	25274	66.13
132	25276	66.89
133	25291	65.82
134	25294	66.56
135	25944	67.17
136	26177	65.83
137	26215	67.11
138	26254	67.07
139	26546	66.34
140	26881	67.17
141	27224	65.17
142	27305	64.37
143	27606	67.28
144	28101	67.11
145	28791	66.16
146	29196	66.53
147	29422	64.94
148	29408	64.98
149	29925	65.01
150	30621	66.06

Index	NSC_Name	Pharmacophore-Fit Score
151	31083	65.75
152	31139	65.4
153	32878	64.91
154	32992	68.78
155	33520	66.37
156	33688	68.54
157	34392	65.99
158	34443	68.04
159	34510	65.99
160	34520	64.72
161	35611	65.49
162	36002	65.66
163	36297	65.33
164	36900	67.88
165	36904	64.97
166	36899	67.88
167	36909	66.26
168	37121	67.51
169	37845	67.65
170	37917	68.09
171	38955	68.46
172	39332	64.73
173	40328	66.54
174	40582	66.1
175	40663	64.62
176	42085	65.6
177	42330	67.36
178	42132	65.6
179	43143	67.36
180	44138	67.43
181	44213	66.38
182	44631	64.09
183	45405	64.78
184	45520	65.92
185	45640	67.91
186	45759	66.79
187	45750	68.81
188	45763	66.15
189	45767	66.53
190	45772	66.35
191	45885	65.39
192	46672	65.29
193	47002	68.81
194	47582	67.46
195	47674	64.27
196	49453	66.39
197	50120	64.86
198	51439	64.78
199	51447	67.36
200	51703	66.93

Index	NSC_Name	Pharmacophore-Fit Score
201	51863	67.58
202	52138	66.98
203	52140	68.79
204	52408	65.48
205	52398	65.98
206	52904	66.42
207	52910	67.36
208	53130	67.65
209	53331	65.79
210	53396	68.57
211	53764	66.32
212	54041	66.57
213	54107	65.52
214	54271	64.61
215	54272	64.62
216	55468	66.51
217	55547	67.37
218	55550	67.15
219	55556	67.34
220	55730	65.21
221	56089	65.62
222	56091	67.39
223	56092	66.31
224	56097	68.5
225	56093	68.01
226	56094	66.75
227	56095	64.92
228	56101	66.63
229	56183	65.95
230	56215	68.83
231	56338	65.16
232	56742	66.59
233	57562	65.21
234	57553	66.93
235	57576	66.99
236	57884	65.9
237	58191	64.6
238	58602	65.88
239	58600	67.01
240	58897	66.49
241	59335	65.76
242	60088	68.57
243	60746	67.92
244	60752	66.85
245	61740	65.43
246	61734	67
247	62387	66.43
248	62386	67.49
249	62513	64.75
250	63832	65.57

Index	NSC_Name	Pharmacophore-Fit Score
251	63946	67.51
252	64614	66.9
253	65380	66.93
254	65881	66.05
255	66058	65.87
256	66057	67.62
257	68123	65.17
258	68182	66.3
259	68228	66.17
260	68858	65.92
261	69191	66.3
262	69305	64.46
263	69335	66.49
264	69308	66.3
265	69435	66.46
266	69436	66.1
267	69864	67.36
268	70311	64.48
269	70380	65.27
270	70544	64.57
271	70731	64.67
272	71359	66.24
273	71427	67.22
274	71424	64.27
275	71429	68.79
276	71426	67.22
277	71428	64.27
278	71854	65.8
279	71903	68.83
280	71953	66.35
281	71973	67.87
282	72271	66.42
283	72393	68.83
284	72961	66.18
285	73316	64.9
286	74447	66.26
287	74505	64.98
288	75340	66.21
289	75355	68.83
290	75357	65.32
291	75813	65.68
292	76179	66.3
293	76282	65.01
294	76347	67.35
295	76352	66.51
296	76988	66.34
297	77381	66.68
298	77486	68.01
299	77593	64.86
300	77683	65.15

Index	NSC_Name	Pharmacophore-Fit Score
301	77685	65.18
302	77963	68.33
303	78769	66.75
304	78911	67.34
305	79039	68.08
306	79073	64.55
307	79156	65.51
308	79200	64.77
309	79219	64.01
310	79217	64.01
311	80653	66.85
312	80800	65.78
313	81164	66.54
314	81513	65.26
315	82037	67.37
316	82043	67.21
317	82047	67.08
318	82050	66.35
319	82398	65.55
320	82635	65.01
321	82825	67.87
322	82907	67.57
323	83053	65.9
324	84415	66.19
325	84519	66.9
326	84629	66.74
327	84630	66.35
328	84628	65.47
329	85465	65.8
330	85525	64.91
331	85783	67.77
332	85998	66.84
333	86777	66.68
334	87518	66.61
335	87517	68.79
336	87859	65.79
337	87913	67.49
338	87909	67.18
339	87914	67.37
340	87915	67.37
341	87916	67.37
342	88860	65.78
343	89203	65.67
344	89223	65.92
345	89270	66.61
346	89613	64.38
347	89697	64.42
348	89874	65.36
349	90420	66.32
350	90421	68.19

Index	NSC_Name	Pharmacophore-Fit Score
351	90506	65.51
352	90978	65.14
353	91001	64.64
354	91303	66.04
355	91443	66.06
356	91537	66.35
357	91551	66.02
358	91698	65.26
359	91862	66.36
360	91863	66.36
361	91865	66.36
362	91866	66.36
363	91868	66.36
364	92094	65.24
365	92173	65.29
366	92192	66.61
367	92172	66.27
368	92309	66.43
369	92431	65.72
370	92737	66.32
371	93066	66.72
372	93068	66.76
373	93419	66.17
374	94023	67.43
375	94100	67.78
376	94735	68.73
377	95113	68.61
378	95103	67.68
379	95099	67.01
380	95100	67.01
381	95945	66.6
382	95946	66.07
383	95944	65.51
384	95957	65.73
385	96318	67.3
386	97195	67.19
387	98670	67.86
388	98697	65.61
389	99193	67.52
390	100037	67.68
391	100053	66.11
392	100879	65.15
393	101035	66.72
394	101166	67.68
395	101165	67.68
396	101589	64.81
397	101612	66.05
398	101742	65.03
399	102241	67.88
400	102254	67.88

Index	NSC_Name	Pharmacophore-Fit Score
401	102868	65.67
402	102907	66.46
403	103062	67.68
404	103106	66.99
405	103561	67.68
406	103798	67.87
407	104794	67.58
408	105019	67.68
409	105398	68.73
410	105405	68.8
411	105828	67.88
412	106046	65.43
413	106047	65.76
414	106128	64.18
415	106393	66.21
416	106446	65.63
417	106486	65.77
418	106489	67.63
419	106501	66.08
420	106539	67.88
421	106541	67.68
422	106538	66.94
423	106555	66.09
424	106675	65.42
425	106770	67.88
426	107000	66.35
427	107505	67.52
428	107506	66.22
429	107711	67.12
430	108116	67
431	108169	66.23
432	108605	67.87
433	109198	67.88
434	109288	65.51
435	109307	67.03
436	109308	67.03
437	109318	66.72
438	109704	67.88
439	110731	65.57
440	110765	64.97
441	111597	67
442	111937	68.76
443	111930	67.06
444	112146	64.95
445	112202	67.61
446	112203	67.43
447	112204	65.93
448	112515	67.69
449	112924	65.25
450	112925	65.29

Index	NSC_Name	Pharmacophore-Fit Score
451	112931	66.37
452	112933	66.3
453	113500	67.71
454	113948	66.07
455	114017	67.77
456	114102	65.39
457	114734	68.17
458	114753	65.19
459	114860	68.14
460	114915	66.23
461	115492	65.18
462	115607	64.96
463	115963	67.25
464	115920	68.73
465	116269	68.72
466	116632	65.69
467	116633	65.69
468	116787	66.6
469	117844	67.44
470	118427	65.34
471	118825	66.3
472	119134	65.46
473	119186	65.87
474	119187	65.92
475	119661	66.3
476	119846	67.52
477	120040	67.73
478	120179	65.69
479	120177	65.69
480	120180	65.69
481	120183	65.69
482	120712	66.57
483	120708	65.24
484	120842	64.96
485	120899	65.02
486	120958	66.51
487	121148	65.69
488	121191	66.22
489	121361	65.6
490	121358	68.06
491	121497	66.3
492	121496	67.54
493	121512	67.57
494	121925	67.88
495	122011	67.31
496	122058	64.09
497	122203	65.41
498	122325	65.76
499	122762	66.73
500	122817	66.15

Index	NSC_Name	Pharmacophore-Fit Score
501	122815	67.27
502	123434	66.51
503	124017	67.34
504	124018	67.34
505	124165	64.2
506	124167	66.06
507	125434	66.37
508	125439	65.15
509	125449	65.98
510	125448	66.35
511	125444	66.08
512	125450	66.17
513	125466	65.75
514	125467	65.55
515	125470	66.23
516	125511	65.82
517	126474	66.22
518	126486	66.14
519	126485	65.97
520	126488	65.61
521	126481	66.03
522	126489	66.24
523	126495	66.22
524	126522	67.77
525	126698	65.79
526	126779	65.4
527	126876	64.98
528	126884	67.86
529	127971	65.93
530	128279	66.52
531	128293	65.47
532	128278	65.63
533	128386	66.29
534	128561	66.51
535	128668	64.28
536	128673	67.88
537	128722	65.31
538	128945	65.69
539	128946	65.69
540	128950	65.69
541	128953	65.7
542	128974	64.45
543	129242	67.53
544	129786	64.65
545	129923	63.69
546	130279	64.7
547	130263	64.88
548	130281	67.69
549	130977	66.48
550	130975	67.31

Index	NSC_Name	Pharmacophore-Fit Score
551	131424	64.23
552	131605	66.14
553	131651	65.51
554	131916	67.88
555	131917	67.88
556	132049	66.68
557	132070	66.38
558	132809	66.32
559	133101	67.22
560	133117	67.92
561	133122	66.48
562	133114	65.93
563	133355	66.22
564	133125	67.04
565	133720	66.5
566	133784	65.11
567	134730	65
568	135704	63.32
569	135710	63.37
570	135754	66.8
571	135756	68
572	136029	67.23
573	136563	67.88
574	136726	64.95
575	136724	67.06
576	136953	65.39
577	136987	64.91
578	137030	65.9
579	137374	65.59
580	137544	66.53
581	137586	66.04
582	137609	67
583	138280	67.88
584	138437	66.57
585	138438	67.61
586	139956	65.96
587	140294	63.34
588	140297	63.37
589	140390	65.4
590	140851	67.34
591	140862	67.88
592	140924	67.14
593	141235	66.51
594	142029	67.88
595	143000	65.97
596	143089	67.27
597	143095	67.51
598	143257	66.32
599	143687	65.6
600	143910	65.02

Index	NSC_Name	Pharmacophore-Fit Score
601	143987	67.43
602	144153	66.95
603	144661	67.68
604	144663	66.03
605	144697	67.65
606	144748	64.17
607	145177	64.64
608	145979	67.38
609	146423	65.93
610	146420	67.2
611	146780	66.3
612	147077	65.38
613	147755	68.78
614	147781	66.73
615	147793	66.02
616	148458	65.9
617	148459	64.95
618	148515	67.69
619	149504	65.04
620	150037	67.88
621	150418	65.8
622	150858	67.26
623	150860	66.3
624	150877	65.03
625	151736	66.3
626	151831	64.34
627	152562	65.81
628	152602	66.42
629	152644	65.58
630	152732	65.72
631	152736	67.32
632	154316	65.73
633	154315	63.4
634	154820	66.07
635	154833	66.07
636	155250	67.87
637	155699	64.95
638	156266	66.63
639	156271	67.48
640	156273	66.01
641	156598	65.45
642	156910	68.83
643	156885	67.54
644	157027	66.8
645	157062	66.17
646	157070	68.12
647	157724	67.87
648	157737	66.51
649	157740	67.88
650	157944	65.78

Index	NSC_Name	Pharmacophore-Fit Score
651	157949	64.89
652	158203	64.2
653	158290	68.78
654	158437	65.41
655	158488	65.95
656	158489	67.87
657	159859	66.07
658	159949	66.52
659	159951	65.98
660	159966	66.27
661	160155	66.94
662	160125	65.57
663	160791	67.52
664	162056	66.66
665	162060	66.66
666	162139	65.49
667	162158	66.7
668	162189	65.71
669	162194	66.68
670	162214	65.71
671	162230	66.12
672	162232	66.68
673	162233	66.09
674	162255	66.07
675	162284	66.68
676	162326	66.05
677	162329	66.68
678	162380	65.54
679	162390	66.05
680	162534	65.13
681	163026	68.79
682	163024	66.92
683	163025	67.05
684	163069	67.29
685	163084	64.6
686	163668	64.63
687	164010	64.28
688	164845	64.53
689	164936	67.47
690	165498	63.98
691	165553	67.87
692	165579	67.14
693	165980	65.08
694	166062	65.66
695	166279	64.74
696	166567	66.05
697	166726	65.27
698	167896	66.51
699	167895	64.99
700	168542	65.56

Index	NSC_Name	Pharmacophore-Fit Score
701	168620	65.59
702	168812	65.74
703	169530	64.82
704	169678	65.37
705	169888	65.66
706	170119	68.04
707	170116	67.39
708	170141	65.65
709	170174	66.34
710	170173	68.27
711	170212	66.43
712	170194	64.79
713	170229	67.14
714	170253	67.89
715	170228	65.94
716	170256	66.58
717	170257	65.72
718	170259	65.65
719	170266	67.62
720	171569	67.88
721	172281	67.23
722	172285	66
723	172303	67.07
724	172599	66.44
725	172626	67.15
726	172843	67.09
727	173937	64.06
728	174187	67.58
729	174120	68.34
730	174793	66.04
731	174876	66.13
732	175153	67.88
733	175374	64.68
734	175379	64.69
735	176135	66.05
736	176979	67.46
737	177254	67.21
738	178248	66.38
739	178836	66.15
740	177001	66.06
741	178838	67.87
742	179176	66.13
743	179460	66.39
744	179470	65.44
745	179471	65.74
746	179475	66.9
747	179473	65.44
748	179652	66.71
749	179624	66.31
750	181926	66.18

Index	NSC_Name	Pharmacophore-Fit Score
751	182862	68.08
752	185309	66.11
753	185316	66.96
754	185324	65.48
755	186309	67.52
756	186308	67.52
757	186914	64.91
758	187684	67.43
759	188491	66.75
760	190656	64.99
761	192746	66.88
762	192745	66.23
763	194865	66.63
764	196553	68.08
765	196554	66.65
766	196590	66.06
767	196838	66.57
768	197191	66.78
769	200711	66.22
770	200710	67.01
771	201338	66.5
772	201339	66.5
773	201462	65.18
774	203336	67.08
775	203448	66.27
776	204856	65.83
777	204855	66.17
778	204984	68.19
779	205375	67.41
780	205374	67.41
781	205377	67.41
782	205378	67.77
783	205376	67.41
784	205380	67.77
785	205379	67.65
786	205381	67.41
787	205559	67.41
788	205558	67.41
789	205567	64.26
790	206189	66.05
791	207110	65.05
792	207117	64.79
793	207873	68.08
794	210513	64.57
795	210898	65.85
796	211154	64.44
797	211274	65.28
798	211340	64.43
799	211434	66.3
800	211721	65.05

Index	NSC_Name	Pharmacophore-Fit Score
801	212010	65.06
802	213271	66.3
803	213269	68.08
804	213274	66.93
805	213273	67.43
806	213287	67.87
807	213276	68.79
808	213272	66.93
809	213280	67.87
810	213275	66.93
811	213284	67.87
812	213286	67.87
813	213288	67.87
814	213289	67.87
815	214024	64.99
816	215556	64.38
817	216751	64.69
818	216755	65.95
819	216758	64.69
820	216765	64.75
821	216772	65.95
822	216774	65.94
823	216780	64.69
824	216787	64.75
825	216788	65.95
826	216795	65.95
827	216956	64.69
828	217067	64.73
829	217084	65.94
830	217346	65.33
831	219981	66.48
832	220040	67.16
833	220470	66.07
834	221018	66.51
835	221099	66.51
836	221098	66.57
837	221659	66.12
838	222356	66.68
839	222365	65.72
840	222363	66.48
841	222366	65.98
842	222367	66.35
843	222358	66.68
844	224095	66.57
845	224278	66.33
846	224279	66.34
847	224293	65.98
848	224294	65.72
849	224297	65.72
850	224302	66

Index	NSC_Name	Pharmacophore-Fit Score
851	224295	65.72
852	224298	66.78
853	224309	66.26
854	225040	67.62
855	225332	68.47
856	226044	67.56
857	226046	68.75
858	226047	66.84
859	226053	67.12
860	226057	68.04
861	226065	66.7
862	226196	66.32
863	226802	68.53
864	226834	67.27
865	226836	67.66
866	226837	67.56
867	226844	67.17
868	226838	65.72
869	226853	66.12
870	226852	66.88
871	226856	66.07
872	226857	67.11
873	226860	66.07
874	226918	66.44
875	226971	67.18
876	227250	66.51
877	227251	66.51
878	227275	65.69
879	227898	67.46
880	231421	64.95
881	231274	66.06
882	231524	64.52
883	231823	67.19
884	231821	66.28
885	231828	68.85
886	231824	65.5
887	231835	66.26
888	231885	67.17
889	231889	65.98
890	231901	66.19
891	231902	66.01
892	231910	65.36
893	231938	66.63
894	232047	66.41
895	232046	65.52
896	232055	67.8
897	232077	65.83
898	232093	64.36
899	233905	66.57
900	234178	67.77

Index	NSC_Name	Pharmacophore-Fit Score
901	234211	66.51
902	234210	66.51
903	234236	67.18
904	234670	66.51
905	234729	66.51
906	235131	65.51
907	237531	66.8
908	236603	66.17
909	236639	66.15
910	238124	65.53
911	238990	67.79
912	239718	66.51
913	240023	66.58
914	240019	66.51
915	240020	66.51
916	240382	65.36
917	240426	67.23
918	240428	66.29
919	242031	65.72
920	243155	66.12
921	243481	65.61
922	243483	65.71
923	243675	66.32
924	243738	66.3
925	244243	67.3
926	244297	66.57
927	244265	68.04
928	244415	66.51
929	244784	66.43
930	244783	66.43
931	245467	65.72
932	246003	65.7
933	246104	65.84
934	246138	66.51
935	246140	67.56
936	246417	67.91
937	248896	67.4
938	248997	65.43
939	249327	66.51
940	249334	67.03
941	250630	66.58
942	251043	68.45
943	251222	66.2
944	251248	66.15
945	252121	65.67
946	252632	66.51
947	252629	66.51
948	252852	66
949	254157	67.96
950	254189	64.65

Index	NSC_Name	Pharmacophore-Fit Score
951	254192	68.09
952	254196	65.62
953	254191	64.65
954	254197	66.03
955	254200	64.65
956	254199	65.71
957	254198	67.27
958	254202	65.24
959	254213	64.61
960	254212	64.91
961	254214	65.23
962	254679	67.79
963	254680	67.79
964	255993	66.33
965	255995	66.3
966	256464	66.57
967	256926	68.79
968	256961	65.23
969	257450	66.37
970	257879	65.24
971	257880	65.25
972	258348	67.5
973	258313	66.64
974	258354	65.16
975	259951	66.05
976	260548	66.51
977	260679	68.62
978	262195	66.51
979	262196	66.57
980	262667	64.86
981	263755	68.04
982	263855	66.33
983	263865	66.51
984	263858	64.99
985	264311	66.51
986	264313	66.51
987	264310	66.51
988	264315	66.57
989	264318	67.31
990	264403	67.41
991	265452	67.78
992	265449	64.93
993	265493	67.93
994	266225	66.51
995	266219	66.67
996	267084	63.74
997	267223	66.51
998	267216	66.67
999	267689	65.28
1,000	268231	66.55

Index	NSC_Name	Pharmacophore-Fit Score
1,001	268233	66.55
1,002	268235	66.51
1,003	268236	66.51
1,004	268709	65.45
1,005	268849	65.73
1,006	268848	66.23
1,007	269137	64.59
1,008	269138	66.57
1,009	269418	65.42
1,010	269741	68.33
1,011	270364	68.7
1,012	270457	65.51
1,013	270519	67.16
1,014	270516	67.87
1,015	270568	65.05
1,016	270752	65.38
1,017	271933	66.51
1,018	272690	65.28
1,019	272679	67.74
1,020	272928	65.7
1,021	274242	67.36
1,022	274245	67.36
1,023	274240	67.36
1,024	274241	67.27
1,025	274244	67.27
1,026	274246	67.36
1,027	274247	67.36
1,028	274248	67.36
1,029	274243	67.36
1,030	274249	65.73
1,031	274250	66.29
1,032	274251	65.75
1,033	274255	66.1
1,034	274538	67.18
1,035	275621	67.16
1,036	275618	66.37
1,037	275625	66.05
1,038	275890	65.67
1,039	276386	66.48
1,040	276391	66.38
1,041	276420	64.72
1,042	277289	66.09
1,043	277567	66.48
1,044	277810	66.51
1,045	277811	66.51
1,046	278178	66.67
1,047	278619	66.81
1,048	278179	66.67
1,049	278334	66.5
1,050	278336	66.5

Index	NSC_Name	Pharmacophore-Fit Score
1,051	278631	65.52
1,052	278633	66.09
1,053	278785	66.33
1,054	279001	65.55
1,055	281313	66.51
1,056	280742	64.67
1,057	281283	67.77
1,058	281310	66.51
1,059	281312	66.51
1,060	281314	66.51
1,061	281311	66.51
1,062	281315	66.51
1,063	281316	66.51
1,064	281372	66.51
1,065	281633	65.51
1,066	282090	66.37
1,067	282178	67.02
1,068	283425	66.55
1,069	283454	66.51
1,070	284180	67.96
1,071	285141	66.51
1,072	286323	66.51
1,073	286617	63.11
1,074	286634	68.07
1,075	287022	66.51
1,076	287076	66.51
1,077	287464	65.68
1,078	287091	65.12
1,079	288388	65.84
1,080	289118	66.51
1,081	289117	66.51
1,082	289490	65.68
1,083	289489	65.88
1,084	289488	65.26
1,085	289564	66.3
1,086	290201	66.51
1,087	290676	64.44
1,088	290678	64.45
1,089	290677	64.44
1,090	290679	64.45
1,091	291098	64.9
1,092	291322	67.75
1,093	291548	66.38
1,094	292135	67.79
1,095	292253	66.11
1,096	293132	65.14
1,097	293539	67.93
1,098	293547	66.92
1,099	293537	67.79
1,100	293542	64.69

Index	NSC_Name	Pharmacophore-Fit Score
1,101	293551	67.79
1,102	293849	65
1,103	293888	65.78
1,104	293891	65.93
1,105	293896	65.78
1,106	294194	66.51
1,107	294399	65.79
1,108	294408	65.31
1,109	294512	66.96
1,110	294945	65.19
1,111	294944	65.18
1,112	294987	66.55
1,113	295272	66.51
1,114	295274	66.51
1,115	295273	66.51
1,116	295275	66.51
1,117	295415	65.27
1,118	295570	66.53
1,119	295571	66.73
1,120	295673	66.3
1,121	295674	66.4
1,122	296341	66.51
1,123	296941	68.78
1,124	296962	66.51
1,125	297558	66.51
1,126	297580	66.51
1,127	298787	66.35
1,128	298882	66.51
1,129	300533	66.31
1,130	301492	66.31
1,131	301745	66.42
1,132	302051	65.21
1,133	302361	65.91
1,134	302383	67.66
1,135	302982	66.34
1,136	302996	65.86
1,137	303793	66.14
1,138	303794	66.19
1,139	304447	66.67
1,140	305723	65.84
1,141	306109	65.63
1,142	305984	66.49
1,143	307994	66.51
1,144	308000	67.54
1,145	309128	65.84
1,146	308875	68.78
1,147	309804	65.94
1,148	310104	65.59
1,149	310834	65.44
1,150	311324	66.42

Index	NSC_Name	Pharmacophore-Fit Score
1,151	311467	67.51
1,152	312810	66.52
1,153	312889	67.97
1,154	312904	66.09
1,155	313416	64.89
1,156	313422	65.46
1,157	314046	65.31
1,158	314048	65.65
1,159	314571	66.48
1,160	315827	65.35
1,161	315835	65.41
1,162	316461	67.87
1,163	317309	65.16
1,164	317312	65.04
1,165	318801	68.09
1,166	318824	66.73
1,167	319044	65.56
1,168	319668	65.94
1,169	319710	65.3
1,170	319758	66.98
1,171	319956	65.97
1,172	320214	66.14
1,173	320215	66.14
1,174	320850	66.14
1,175	322095	67.87
1,176	322096	67.87
1,177	322099	66.07
1,178	323743	64.1
1,179	324326	66.48
1,180	324361	65.4
1,181	325014	65.75
1,182	326924	68.21
1,183	327189	66.97
1,184	327693	66.14
1,185	327708	65.47
1,186	327926	66.51
1,187	327924	67.79
1,188	328399	67.08
1,189	329055	65.79
1,190	329238	66.51
1,191	329341	66.03
1,192	330758	67.3
1,193	330788	66.48
1,194	331616	66.57
1,195	331778	67.18
1,196	331916	64.91
1,197	332677	65.52
1,198	333184	65.27
1,199	333530	68.78
1,200	334045	66.57

Index	NSC_Name	Pharmacophore-Fit Score
1,201	334213	66.68
1,202	334716	65.26
1,203	335758	65.08
1,204	335998	66.45
1,205	335994	66.51
1,206	336242	66.51
1,207	337234	66.57
1,208	337235	66.56
1,209	337237	67.43
1,210	337721	65.52
1,211	337761	64.67
1,212	337776	64.96
1,213	337775	66.57
1,214	338033	66.42
1,215	338187	66.05
1,216	338240	67.52
1,217	338659	66.39
1,218	338660	66.37
1,219	338943	66.59
1,220	338953	66.91
1,221	339175	65.67
1,222	339187	67.89
1,223	339699	66.06
1,224	340051	68.76
1,225	340065	66.63
1,226	340847	67.83
1,227	342982	66.2
1,228	342673	66.63
1,229	343033	64.27
1,230	343957	66.12
1,231	344263	67.17
1,232	344512	66.57
1,233	344507	66.51
1,234	344682	66.57
1,235	345142	67.26
1,236	345413	67.66
1,237	345852	65.28
1,238	347471	66.51
1,239	347784	66.53
1,240	347786	67.77
1,241	348102	66.25
1,242	348694	66.58
1,243	349461	66.14
1,244	349631	68.07
1,245	351130	65.67
1,246	351159	65.35
1,247	351376	66.2
1,248	351576	66.75
1,249	351572	65.7
1,250	351748	66.48

Index	NSC_Name	Pharmacophore-Fit Score
1,251	352375	67.87
1,252	352721	63.26
1,253	352890	66.48
1,254	353094	63.12
1,255	353459	66.53
1,256	354088	66.42
1,257	354475	65.61
1,258	354627	67.77
1,259	354955	64.92
1,260	354958	64.91
1,261	354956	64.91
1,262	356555	64.74
1,263	356802	66.49
1,264	360639	66.45
1,265	360637	66.51
1,266	360640	66.57
1,267	360645	66.57
1,268	360641	66.57
1,269	361412	65.74
1,270	361755	65.42
1,271	361756	65.39
1,272	361804	64.43
1,273	362402	67.19
1,274	363218	66.51
1,275	363223	66.48
1,276	363425	66.51
1,277	363818	66.46
1,278	363986	66.31
1,279	365412	68.08
1,280	365411	68.08
1,281	365455	65.3
1,282	365623	65.88
1,283	366752	65.13
1,284	366745	66.74
1,285	367090	65.18
1,286	367117	66.76
1,287	367264	66.65
1,288	368010	66.09
1,289	368667	68.79
1,290	368715	66.76
1,291	369056	65.61
1,292	369002	66.93
1,293	369313	64.77
1,294	369758	67.98
1,295	370149	66.13
1,296	370273	66.74
1,297	370271	66.74
1,298	370274	66.74
1,299	370270	66.92
1,300	370346	64.9

Index	NSC_Name	Pharmacophore-Fit Score
1,301	370380	67.77
1,302	370383	67.56
1,303	370471	67.79
1,304	370472	67.79
1,305	370605	67.93
1,306	371040	66.28
1,307	371056	66.65
1,308	371800	67.27
1,309	371792	66.51
1,310	371798	67.78
1,311	371838	66.48
1,312	372329	66.18
1,313	372327	65.57
1,314	372326	66.3
1,315	373816	66.82
1,316	374036	64.44
1,317	374118	68.79
1,318	374900	67.87
1,319	374529	66.51
1,320	374531	66.51
1,321	376269	63.99
1,322	376971	66.55
1,323	377155	65.13
1,324	377613	68.03
1,325	379419	66.08
1,326	380531	66.39
1,327	381848	65.29
1,328	381850	65.29
1,329	381866	67.27
1,330	382683	66.91
1,331	382773	67.41
1,332	382873	66.89
1,333	383033	67.23
1,334	383123	67.11
1,335	383465	66.22
1,336	400027	64.6
1,337	400145	66.35
1,338	400036	64.6
1,339	400275	67.59
1,340	400360	66.23
1,341	400443	66.41
1,342	400525	66.37
1,343	400754	65.8
1,344	400963	64.64
1,345	401114	65.85
1,346	401403	67.29
1,347	401605	66.1
1,348	401822	67.04
1,349	401824	66.8
1,350	401847	64.85

Index	NSC_Name	Pharmacophore-Fit Score
1,351	401843	66.5
1,352	401848	64.63
1,353	401860	66.54
1,354	402078	66.12
1,355	402483	68.08
1,356	402484	67.42
1,357	402486	67.32
1,358	402666	66.46
1,359	402665	65.06
1,360	402668	66.92
1,361	403092	68.14
1,362	403178	68.79
1,363	403187	67.04
1,364	403351	67.39
1,365	403468	65.8
1,366	403465	67.43
1,367	403487	65.27
1,368	403946	65.55
1,369	403945	68.71
1,370	403947	68.19
1,371	403948	67.15
1,372	403950	67.58
1,373	403955	66.88
1,374	404264	67.65
1,375	404566	66.38
1,376	404560	66.31
1,377	405284	64.78
1,378	405276	67.88
1,379	405666	67.52
1,380	405929	66.21
1,381	405936	66.59
1,382	405937	66.01
1,383	406239	67.64
1,384	406434	65.09
1,385	406436	65.09
1,386	406435	65.09
1,387	406891	65.62
1,388	406942	68.78
1,389	407108	66.55
1,390	407185	67.88
1,391	407283	67.05
1,392	407297	65.94
1,393	407298	64.87
1,394	407348	64.85
1,395	407338	66.75
1,396	407809	67.58
1,397	408105	65.46
1,398	408121	67.31
1,399	408123	67.31
1,400	408122	67.31

Index	NSC_Name	Pharmacophore-Fit Score
1,401	408130	67.31
1,402	408361	67.17
1,403	408372	67.68
1,404	409166	66.02
1,405	409729	67.06
1,406	409823	67.88
1,407	409925	65.89
1,408	506863	63.86
1,409	507922	66.38
1,410	508287	68.12
1,411	509568	67.43
1,412	513473	68.8
1,413	515360	67.66
1,414	516183	67.66
1,415	517192	67.68
1,416	517467	66.63
1,417	519708	65.88
1,418	519980	65.6
1,419	520474	67.36
1,420	520703	67.02
1,421	521508	66.07
1,422	521466	66.68
1,423	521703	66.53
1,424	521711	67.37
1,425	521713	67.27
1,426	521712	67.27
1,427	521714	67.27
1,428	521716	67.76
1,429	521793	67.36
1,430	521957	66.89
1,431	522484	66.57
1,432	522573	66.51
1,433	522704	67.73
1,434	523981	66.07
1,435	524244	65.23
1,436	524272	66.07
1,437	524462	65.69
1,438	524758	66.07
1,439	525604	68.04
1,440	526374	67.27
1,441	526376	67.27
1,442	526476	66.87
1,443	526520	65.42
1,444	527033	66.3
1,445	527336	65.41
1,446	528018	64.91
1,447	528169	66.32
1,448	529855	67.04
1,449	600271	64.11
1,450	600275	63.99

Index	NSC_Name	Pharmacophore-Fit Score
1,451	600274	64.11
1,452	600779	65.96
1,453	600413	66.65
1,454	601516	67.31
1,455	602356	67.05
1,456	602355	66.99
1,457	602614	65.24
1,458	601346	66.48
1,459	602630	64.61
1,460	602743	67.35
1,461	602881	64.2
1,462	602892	65.27
1,463	602913	66.51
1,464	602896	68.78
1,465	603972	67.43
1,466	603831	65.54
1,467	604985	67.41
1,468	605735	66.27
1,469	605737	66.27
1,470	605736	66.42
1,471	605739	66.27
1,472	605740	66.42
1,473	605741	66.27
1,474	605704	66.46
1,475	606253	65.07
1,476	606254	65.05
1,477	606255	68.79
1,478	606256	66.94
1,479	606257	66.93
1,480	606258	65.91
1,481	606402	68.78
1,482	606403	67.43
1,483	607151	68.32
1,484	607156	68.5
1,485	607157	68.52
1,486	607160	66.97
1,487	607154	68.36
1,488	607158	68.32
1,489	607159	68.52
1,490	607568	64.93
1,491	607607	64.81
1,492	607635	67.05
1,493	607634	66.44
1,494	607864	67.07
1,495	607865	67.07
1,496	608056	67.09
1,497	608642	66.42
1,498	608646	66.42
1,499	608647	66.27
1,500	608645	66.27

Index	NSC_Name	Pharmacophore-Fit Score
1,501	608649	66.42
1,502	608653	66.42
1,503	608652	66.27
1,504	608654	66.27
1,505	609240	66.42
1,506	609244	66.42
1,507	609245	66.27
1,508	609516	65.19
1,509	609518	67.95
1,510	609517	65.52
1,511	609519	65.19
1,512	609520	67.95
1,513	609594	66.46
1,514	609595	67.46
1,515	609597	67.6
1,516	610183	64.33
1,517	610331	67.51
1,518	610927	66.48
1,519	610966	66.49
1,520	611422	67.88
1,521	611433	67.88
1,522	611893	65.05
1,523	611895	65.91
1,524	611897	65.95
1,525	611894	63.85
1,526	611913	66.02
1,527	612223	66.25
1,528	612236	68.56
1,529	612488	66.59
1,530	612866	66.75
1,531	612955	66.61
1,532	614379	66.38
1,533	614438	65.19
1,534	614510	65.24
1,535	614507	65.85
1,536	614610	66.3
1,537	614833	66.56
1,538	614904	68.8
1,539	615479	64.69
1,540	615594	66.71
1,541	616809	66.19
1,542	616960	68.78
1,543	616982	66.04
1,544	617744	66.51
1,545	617794	68.8
1,546	617745	67.15
1,547	617792	65.69
1,548	617795	68.8
1,549	617793	66.52
1,550	618214	65.31

Index	NSC_Name	Pharmacophore-Fit Score
1,551	618213	65.31
1,552	618270	66.38
1,553	618454	67.1
1,554	618462	66.61
1,555	618490	65.32
1,556	618493	65.7
1,557	618522	65.54
1,558	618646	64.22
1,559	619797	66.38
1,560	619900	64.51
1,561	619899	66.81
1,562	620289	64.64
1,563	620321	66.87
1,564	620320	67.86
1,565	620635	67.87
1,566	620673	68.52
1,567	620680	65.88
1,568	620686	67.52
1,569	620690	68.46
1,570	620695	65.05
1,571	621178	67.77
1,572	621189	65.37
1,573	621608	65.9
1,574	621684	68.71
1,575	622559	67.66
1,576	622472	68.12
1,577	622474	67.43
1,578	622638	66.3
1,579	622971	66.38
1,580	622982	67.43
1,581	623667	66.51
1,582	623906	64.43
1,583	624018	67.02
1,584	624019	66.96
1,585	624415	64.42
1,586	624416	64.43
1,587	624420	64.43
1,588	624419	64.43
1,589	624423	64.43
1,590	624422	64.43
1,591	624435	64.43
1,592	624513	65.37
1,593	624647	65.72
1,594	624649	65.79
1,595	624646	68.36
1,596	624648	68.32
1,597	624650	68.32
1,598	624652	68.31
1,599	624654	68.5
1,600	624808	66.28

Index	NSC_Name	Pharmacophore-Fit Score
1,601	625739	66.57
1,602	625821	65.81
1,603	625823	65.81
1,604	625822	64.73
1,605	625824	65.81
1,606	625825	65.81
1,607	626094	66.84
1,608	625827	64.75
1,609	625835	65.75
1,610	625836	66.93
1,611	626966	65.07
1,612	627046	66.92
1,613	627044	66.46
1,614	627085	66.64
1,615	627525	66.3
1,616	627526	66.3
1,617	627551	67.21
1,618	627780	65.49
1,619	628050	66.06
1,620	628083	68.79
1,621	628412	66.71
1,622	628413	66.71
1,623	628415	66.71
1,624	628416	66.71
1,625	628417	66.71
1,626	628414	66.71
1,627	629050	66.48
1,628	629114	68.8
1,629	629117	68.79
1,630	629380	67.01
1,631	630254	67.17
1,632	630249	66.88
1,633	631176	65.67
1,634	631262	68.12
1,635	631406	66.69
1,636	631600	65.58
1,637	631814	66.1
1,638	631846	66.1
1,639	631848	65.3
1,640	631852	65.94
1,641	631951	66.3
1,642	631952	66.3
1,643	632116	65.8
1,644	632822	66.18
1,645	632826	66.18
1,646	632828	66.18
1,647	632827	66.18
1,648	632823	66.18
1,649	632829	66.18
1,650	632901	66.34

Index	NSC_Name	Pharmacophore-Fit Score
1,651	633357	68.78
1,652	633358	68.78
1,653	633362	66.75
1,654	633359	68.78
1,655	633360	68.78
1,656	633364	66.75
1,657	633363	68.78
1,658	633366	68.78
1,659	633367	68.78
1,660	634665	65.81
1,661	634679	66.93
1,662	634803	68.78
1,663	635401	67.19
1,664	635419	65.66
1,665	635422	65.65
1,666	635455	66.99
1,667	635454	67.02
1,668	635532	65.81
1,669	635540	64.73
1,670	635549	65.8
1,671	635555	66.22
1,672	635558	65.9
1,673	636186	64.44
1,674	636227	65.54
1,675	636380	65.91
1,676	636592	67.32
1,677	636776	67.74
1,678	636794	65
1,679	636796	65
1,680	636984	65
1,681	637520	65.98
1,682	637525	65.25
1,683	637670	66.65
1,684	638024	67.87
1,685	638030	66.51
1,686	638130	67.87
1,687	638131	67.87
1,688	638240	65.35
1,689	639659	65.81
1,690	640073	67.82
1,691	640074	67.56
1,692	640199	68.04
1,693	640156	67.19
1,694	640463	66.61
1,695	640468	65.51
1,696	640525	66.84
1,697	640749	66.49
1,698	640972	65.05
1,699	641130	68.78
1,700	641136	68.78

Index	NSC_Name	Pharmacophore-Fit Score
1,701	641251	65.75
1,702	641295	66.62
1,703	641403	66.33
1,704	641413	65.74
1,705	641598	66.62
1,706	641602	67.29
1,707	641603	67.29
1,708	641617	67.01
1,709	641616	65.28
1,710	641613	66.49
1,711	641713	65.9
1,712	641816	66.62
1,713	641818	66.25
1,714	641996	66.29
1,715	642010	65.83
1,716	642019	65.75
1,717	642029	65.49
1,718	642049	66.53
1,719	642055	65.38
1,720	642043	66.2
1,721	642638	66.44
1,722	642639	66.62
1,723	642647	66.59
1,724	643003	68.78
1,725	643788	66.07
1,726	643749	68.78
1,727	643789	66.07
1,728	643790	66.07
1,729	643791	66.07
1,730	644015	68.55
1,731	644728	65.02
1,732	644731	67.28
1,733	644780	65.75
1,734	644784	65.75
1,735	644782	65.38
1,736	644793	66.54
1,737	644861	66.14
1,738	645748	67.52
1,739	645746	65.9
1,740	645975	67.88
1,741	645990	65.31
1,742	646087	66.07
1,743	646471	66.65
1,744	646613	66.78
1,745	646818	65.75
1,746	646815	65.68
1,747	646820	66.07
1,748	646828	65.03
1,749	646834	67.28
1,750	647089	65.69

Index	NSC_Name	Pharmacophore-Fit Score
1,751	647529	66.06
1,752	647576	67.24
1,753	647574	65.93
1,754	647586	67.28
1,755	647583	67.28
1,756	647591	67.28
1,757	647625	66.18
1,758	648500	67.91
1,759	648606	65.6
1,760	648610	65.68
1,761	648636	65.68
1,762	648652	66.02
1,763	648764	68.78
1,764	648762	67.72
1,765	649058	66.8
1,766	649161	65.12
1,767	649162	65.12
1,768	649172	67
1,769	649413	68.77
1,770	649646	64.94
1,771	649769	64.62
1,772	649800	65.5
1,773	649837	64.35
1,774	650426	66.98
1,775	650561	66.36
1,776	650721	64.8
1,777	650993	63.13
1,778	651412	66.44
1,779	651688	66.26
1,780	651734	66.65
1,781	651738	65.65
1,782	651808	66.57
1,783	651813	68.78
1,784	652620	67.88
1,785	654102	66.21
1,786	654223	66.52
1,787	655156	66.3
1,788	655649	68.14
1,789	655893	67.31
1,790	656346	68.14
1,791	656348	68.14
1,792	656349	68.14
1,793	656347	68.14
1,794	656351	67.88
1,795	656352	67.88
1,796	656362	67.96
1,797	656643	65.84
1,798	656907	64.82
1,799	657994	65.27
1,800	658885	65.38

Index	NSC_Name	Pharmacophore-Fit Score
1,801	659182	64.8
1,802	659186	65.6
1,803	659194	67.18
1,804	659195	65.68
1,805	659181	66.33
1,806	659211	67.89
1,807	659215	64.84
1,808	659209	67.1
1,809	659217	66.59
1,810	659216	65.98
1,811	659220	65.79
1,812	659226	67.17
1,813	659227	65.95
1,814	659229	67.28
1,815	659255	65.95
1,816	659258	66.24
1,817	659266	65.22
1,818	659268	65.49
1,819	659225	65.64
1,820	660033	67.13
1,821	660034	65.85
1,822	660245	66.31
1,823	660304	65.05
1,824	660305	66.68
1,825	660306	65.54
1,826	660303	66.9
1,827	660307	66.86
1,828	661460	66.3
1,829	661464	65.75
1,830	661803	64.59
1,831	663006	67.58
1,832	663005	67.58
1,833	663881	65.47
1,834	663880	67.14
1,835	663882	67.14
1,836	663990	67.17
1,837	663999	65.68
1,838	664182	66.65
1,839	664186	66.68
1,840	664272	65.78
1,841	664290	65.98
1,842	664326	66.08
1,843	664335	65.53
1,844	664339	66.9
1,845	664402	66.18
1,846	664703	66.63
1,847	664907	67.58
1,848	665510	68.52
1,849	665513	68.52
1,850	665514	64.13

Index	NSC_Name	Pharmacophore-Fit Score
1,851	665511	68.79
1,852	665515	64.56
1,853	665517	67
1,854	665516	67
1,855	665904	65.82
1,856	666125	66.29
1,857	666129	64.75
1,858	666175	65.87
1,859	666211	65.97
1,860	666271	65.21
1,861	666982	67.15
1,862	667543	67.33
1,863	667649	66.45
1,864	667647	68.79
1,865	667653	68.73
1,866	667735	66.06
1,867	668051	64.36
1,868	668415	65.45
1,869	668476	66.25
1,870	669027	64.87
1,871	669466	65.22
1,872	670000	66.68
1,873	670020	68.09
1,874	671266	68.71
1,875	671297	66.98
1,876	671301	66.39
1,877	671928	67.92
1,878	672076	67.56
1,879	673336	65.62
1,880	673353	68.78
1,881	673352	68.78
1,882	673354	68.78
1,883	673418	65.73
1,884	674008	64.95
1,885	674050	67.21
1,886	674051	65.27
1,887	674184	66
1,888	674317	66.7
1,889	674624	68.78
1,890	674626	66.73
1,891	674627	66.84
1,892	674628	68.78
1,893	675810	64.58
1,894	676388	66.35
1,895	676416	65.4
1,896	676445	66.31
1,897	676453	65.24
1,898	676455	64.68
1,899	676463	65.86
1,900	676618	68.16

Index	NSC_Name	Pharmacophore-Fit Score
1,901	676677	65.79
1,902	676689	67.87
1,903	676770	66.85
1,904	676780	66
1,905	676789	65.33
1,906	676797	65.89
1,907	676775	66.19
1,908	676777	66.19
1,909	676790	65.95
1,910	676979	65.86
1,911	677219	67.03
1,912	677500	66.53
1,913	677665	65.66
1,914	677673	68.16
1,915	677933	66.3
1,916	677934	65.43
1,917	677995	65.88
1,918	677996	65.88
1,919	678085	67.27
1,920	678101	66.34
1,921	679210	66.38
1,922	679207	66.54
1,923	679216	66.38
1,924	679211	66.54
1,925	680103	67.49
1,926	680105	67.49
1,927	680108	66.29
1,928	680115	67.72
1,929	680116	67.72
1,930	680117	67.72
1,931	681021	65.85
1,932	681227	66.91
1,933	681226	65.54
1,934	681231	64.93
1,935	682345	66.15
1,936	682755	64.22
1,937	683132	65.94
1,938	683367	64.94
1,939	683580	68.72
1,940	683596	64.91
1,941	683597	65.03
1,942	683693	64.87
1,943	684576	66.3
1,944	684579	65.88
1,945	684577	66.3
1,946	684912	67.24
1,947	685265	67.03
1,948	685266	67.03
1,949	685268	67.03
1,950	685273	68.53

Index	NSC_Name	Pharmacophore-Fit Score
1,951	685274	68.53
1,952	685276	68.53
1,953	685275	67.55
1,954	685277	66.47
1,955	685278	66.47
1,956	685279	66.47
1,957	685280	66.47
1,958	685672	66.91
1,959	686040	68.45
1,960	686041	68.45
1,961	685993	66.06
1,962	685994	66.06
1,963	686352	67.87
1,964	686499	68.75
1,965	686651	65.04
1,966	686656	64.9
1,967	686665	64.91
1,968	686669	64.62
1,969	686781	66.23
1,970	686782	66.23
1,971	687509	65.93
1,972	687510	65.93
1,973	688076	66.96
1,974	688227	68.72
1,975	688958	67.19
1,976	689861	64.94
1,977	690199	66.99
1,978	690211	67.38
1,979	690212	65.17
1,980	691092	68.1
1,981	691095	67.87
1,982	691099	65.74
1,983	691436	67.07
1,984	691615	66.14
1,985	692287	67.23
1,986	692295	67.23
1,987	692598	66
1,988	692636	64.82
1,989	693220	68.1
1,990	693219	68.1
1,991	694280	64.91
1,992	694346	65.51
1,993	694923	65.05
1,994	695252	68.74
1,995	695569	65.21
1,996	695568	65.21
1,997	695663	64.66
1,998	695831	65.63
1,999	696076	64.6
2,000	696224	66.69

Index	NSC_Name	Pharmacophore-Fit Score
2,001	696899	64.35
2,002	697143	66.22
2,003	697142	66.2
2,004	697276	66.54
2,005	697274	66.94
2,006	697730	66.62
2,007	698030	65.36
2,008	698989	64.63
2,009	698990	64.65
2,010	698993	64.64
2,011	698996	64.94
2,012	698999	64.94
2,013	699024	68.16
2,014	699025	67.78
2,015	699188	68.24
2,016	699466	65.6
2,017	699854	64.99
2,018	699853	65
2,019	700145	68.78
2,020	700711	64.9
2,021	700706	64.92
2,022	700709	64.92
2,023	702323	63.75
2,024	702689	64.64
2,025	703213	68.78
2,026	704235	65.88
2,027	704329	65.3
2,028	704330	65.3
2,029	704331	65.3
2,030	704332	65.3
2,031	704558	66.87
2,032	704859	68.78
2,033	704860	68.64
2,034	704861	68.64
2,035	704862	65.33
2,036	704863	65.33
2,037	705951	67.39
2,038	705943	67.88
2,039	706195	68.33
2,040	706387	65.09
2,041	706388	65.05
2,042	706804	68.73
2,043	706839	64.18
2,044	706965	66.7
2,045	707055	66.79
2,046	707231	66.45
2,047	707409	66.28
2,048	707565	64.1
2,049	707566	64.1
2,050	707564	65.34

Index	NSC_Name	Pharmacophore-Fit Score
2,051	707568	64.1
2,052	707570	64.1
2,053	707571	64.1
2,054	707572	64.1
2,055	707800	66.52
2,056	708433	66.03
2,057	708473	66.11
2,058	708475	66.64
2,059	708496	68.71
2,060	708828	66.52
2,061	708928	66.88
2,062	708927	65.76
2,063	708926	68.7
2,064	709343	64.39
2,065	709455	65.19
2,066	709456	65.12
2,067	709457	65.19
2,068	709459	65.12
2,069	709460	65.18
2,070	709462	65.19
2,071	709458	65.21
2,072	709464	65.21
2,073	709461	65.19
2,074	709463	65.19
2,075	709467	65.21
2,076	709466	65.19
2,077	709465	65.21
2,078	709381	65.98
2,079	709862	66.3
2,080	709865	66.3
2,081	709863	66.31
2,082	709901	65.48
2,083	709882	66.94
2,084	710092	65.18
2,085	709884	65.85
2,086	710091	68.78
2,087	710167	65.96
2,088	710090	66.94
2,089	710093	65.18
2,090	710288	63.72
2,091	710353	68.78
2,092	710354	68.78
2,093	710355	68.78
2,094	711181	67.79
2,095	711182	68.12
2,096	711715	64.1
2,097	711714	64.84
2,098	711711	64.09
2,099	712803	68.78
2,100	713564	67.43

Index	NSC_Name	Pharmacophore-Fit Score
2,101	713771	65.86
2,102	714417	64.33
2,103	714601	66.94
2,104	714606	68.78
2,105	714603	66.05
2,106	714607	66.94
2,107	714609	65.85
2,108	714605	66.94
2,109	714639	63.09
2,110	714604	67.02
2,111	714602	65.85
2,112	715238	67.99
2,113	715237	68.74
2,114	715239	68
2,115	716021	68.79
2,116	716028	68.79
2,117	716029	68.79
2,118	716032	67.79
2,119	716140	65
2,120	716231	66.84
2,121	716515	65.52
2,122	716961	68.23
2,123	716975	67.64
2,124	717098	66.58
2,125	717503	66.6
2,126	717505	66.6
2,127	718005	68.75
2,128	717859	68.78
2,129	717862	66.31
2,130	718012	65.77
2,131	719159	68.42
2,132	719924	63.83
2,133	720215	66.25
2,134	720212	67.07
2,135	720216	67.4
2,136	720313	66.19
2,137	720314	66.82
2,138	721048	66.63
2,139	721382	66.43
2,140	722438	63.93
2,141	722450	66.51
2,142	722451	66.49
2,143	724109	66.66
2,144	724520	66.25
2,145	724525	66.26
2,146	724805	65.72
2,147	725321	66.23
2,148	725699	65.75
2,149	725817	66.32
2,150	726105	66.89

Index	NSC_Name	Pharmacophore-Fit Score
2,151	726553	65.31
2,152	727561	66.34
2,153	727562	66.34
2,154	727709	63.74
2,155	727908	67.18
2,156	727911	66.11
2,157	727910	66.12
2,158	728338	64.03
2,159	728339	64.39
2,160	728914	65.34
2,161	728915	64.95
2,162	729643	68.27
2,163	729646	66.96
2,164	729647	66.96
2,165	729651	66.32
2,166	729868	66.19
2,167	729874	66.29
2,168	730058	66.83
2,169	730061	65.27
2,170	730064	66.6
2,171	730062	68.18
2,172	730065	67.88
2,173	730180	66.3
2,174	730176	66.16
2,175	731174	68.45
2,176	731523	66.33
2,177	731924	67.12
2,178	732254	68.45
2,179	732261	68.45
2,180	732271	68.45
2,181	732275	68.45
2,182	733428	67.31
2,183	733427	65.64
2,184	733760	65.26
2,185	734000	66.24
2,186	734001	66.24
2,187	734044	66.53
2,188	734045	66.53
2,189	734046	66.53
2,190	734066	65.84
2,191	734266	68.37
2,192	734268	65.31
2,193	734269	65.23
2,194	734791	65.73
2,195	735809	68.84
2,196	736351	63.06
2,197	736921	65.87
2,198	736919	65.84
2,199	736922	65.84
2,200	737082	66.16

Index	NSC_Name	Pharmacophore-Fit Score
2,201	737142	65.03
2,202	741078	66.3
2,203	741197	65.84
2,204	741682	66.38
2,205	742190	66.8
2,206	742189	65.89
2,207	742191	66.14
2,208	742348	65.9
2,209	742410	66.69
2,210	742480	66.42
2,211	742838	67.88
2,212	742886	66.89
2,213	742888	66.89
2,214	742887	65.65
2,215	742894	65.08
2,216	743229	65.57
2,217	744263	66.61
2,218	744358	67.02
2,219	744462	64.75
2,220	744912	65.53
2,221	745502	65.16
2,222	746271	66.47
2,223	746351	64.65
2,224	746616	66.7
2,225	746930	64.48
2,226	746933	65.64
2,227	746928	66.7
2,228	746931	64.51
2,229	747135	65.12
2,230	748122	63.46
2,231	748126	63.38
2,232	748416	65.66
2,233	748727	66.3

4.3 PREPARING THE RECEPTOR:

For docking studies, X-ray crystallographic structure of galectin-1 protein with higher resolution was downloaded from protein data bank e.g. PDB ID 1W6O & resolution1.90 A^0. Gal-1 is homodimer with chain A & B and bulky protein. So we selected gal-1 chain A and removed chain B using chimera software. Then protein was given as input file in Gromacs for energy minimization and simulation. We applied Gromacs96 43a1 force field and simulated in water environment. Total 5156 water molecules were added to the system. We used steepest descent algorithm with tolerance (Fmax) 1.00000e+03 and number of steps 50000.Our protein's lowest energy coordinates were generated with steepest descents converged to Fmax < 1000 in 249 steps. And 249 steps have completed in 248.00 ps with

-244279 kcal/mol average potential energy (Figure 7).

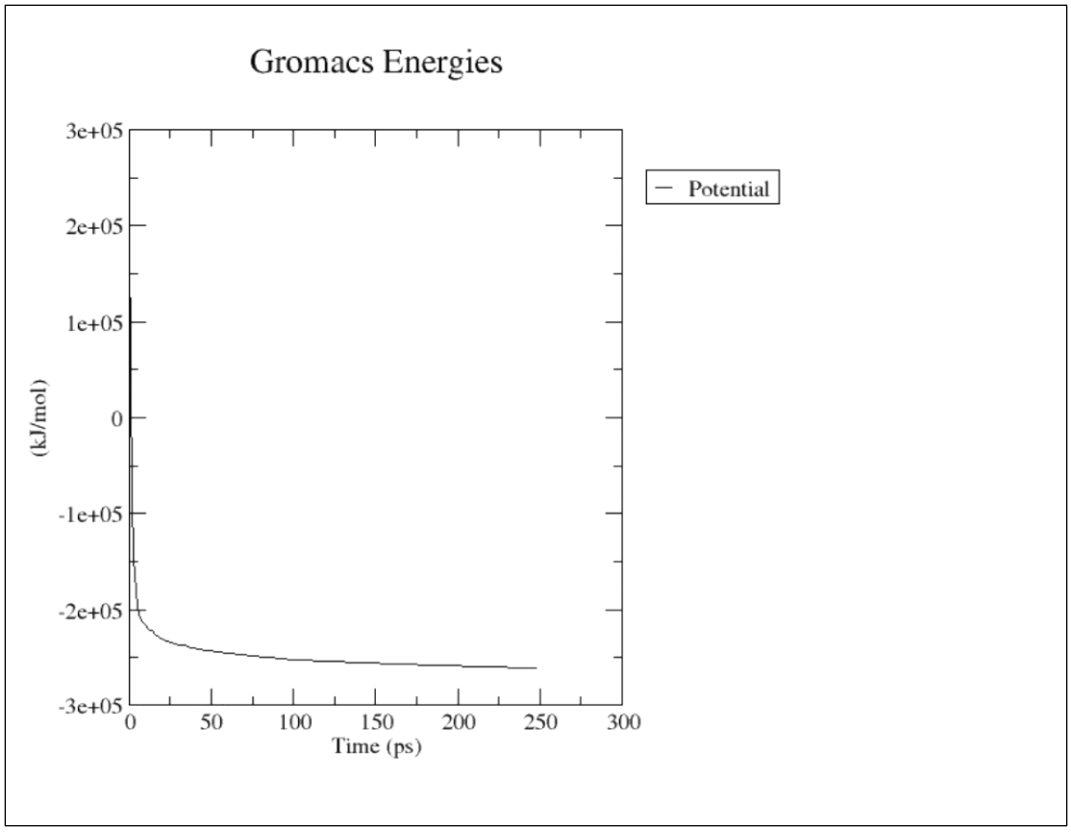


Figure 7. Protein energy minimization using Gromacs v 4.5.6

4.4 BINDING SITE PREDICTION:

Binding sites were obtained from literature and also predicted from NCBI website. After submitting PDB ID 1W6O in NCBI, we found sugar binding sites residues that also mention as carbohydrate recognition domain (CRD) residues in literature. These conserved domain residues are His-44, Asn-46, Arg-48, His-52, Asn-61, Trp-68, Glu-71, and Arg-73 (Figure 8).

NCBI result of 1W6O.

Chain A, X-Ray Crystal Structure Of C2s Human Galectin-1 complexed with Lactose

ORIGIN

1 asglvasnln lkpgexlrvr gevapdaksf vlnlgkdsnn lclhfnprfn ahgdantivc
61 nskddgawgt eqreavfpfq pgsvaevcit fdqanltvkl pdgyefkfpn rlnleainym
121 aadgdfkikc vafd

//

order(44,46,48,59,61,68,71,73)
/site_type="other"
/note="sugar binding pocket [chemical binding]"
/db_xref="CDD:238025"

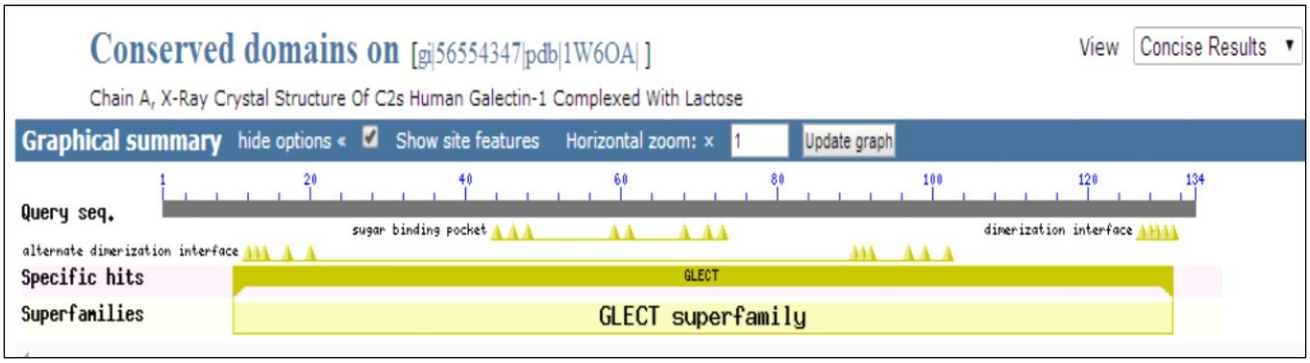


Figure 8. Conserved domain on Human Galectin-1 Complexed with Lactose (1W6O)

(http://www.ncbi.nlm.nih.gov/)

4.5 MOLECULAR DOCKING:

Molecular docking was performed to observe the small molecular affinity and interaction with CRD residue. All database hits were docked into identified carbohydrate recognition domain, using automated Autodock_Vina program. Molecular docking results were used to post- docking filter to select the compounds those binding energy was ≥-7.0Kcal/mol.316 compounds out of 2233 docked hits were filtered and used for further analysis (Table-3).

4.6 DOCKING ANALYSIS:

MGLTools v1.5.6 provide autodock vina results analysis program. The graphical presentation of docked confirmation of compounds were analysed for all 20 clusters. The clustering define interaction between active site residues of macromolecule and ligand. Based on number of cluster, docked results were further filtered. Those compound had ≥13 clusters out of 20 were selected. Finally, those compounds with binding affinity ≥-8.0Kcal/mol and cluster ≥ 13 considered as best compounds. We found 6 compounds, those fulfilled the both parameters are NCI_HITS 50, 379, 592, 1713, 1938 & 1992 (Table-6). In LigPlot+ analysis, 6 best compounds named NSC4320, NSC95099, NSC140924, NSC641818, NSC683367, and NSC694346 were shown H-bond and hydrophobic interaction with CRD residues : His-44, Asn-46, Arg-48, His-52, Asn-61, Trp-68, Glu-71, and Arg-73 (Figure 9).

Table: 3 Binding energy calculation of hits (≥-7.0Kcal/mol)

SL No.	NCI_Hits	Best affinity (Kcal/mol)
1	24	-7.7
2	34	-7.1
3	43	-7.2
4	50	-8
5	53	-7.4
6	54	-7.6
7	56	-7.5
8	63	-7.5
9	64	-7.8
10	65	-7.2
11	66	-7.4
12	112	-7.1
13	115	-7.4
14	123	-7.2
15	125	-7.6
16	126	-7.5
17	161	-7.3
18	162	-8.4
19	183	-7
20	192	-7.4
21	193	-7
22	203	-7.2
23	209	-7.3
24	210	-7.6
25	228	-7.6
26	263	-7.3
27	269	-7.1
28	314	-7.6
29	329	-7.5
30	340	-7.4
31	341	-7.3
32	357	-7.8
33	366	-7.4
34	367	-7
35	376	-7.8
36	379	-8.1
37	380	-7.3
38	395	-7.2
39	407	-8.7
40	416	-7.7
41	423	-7.3
42	438	-7.3
43	464	-8.1
44	466	-7
45	468	-7.2
46	489	-7
47	532	-7
48	533	-7.5
49	545	-7.8
50	553	-7.3

SL No.	NCI_Hits	Best affinity (Kcal/mol)
51	577	-8.1
52	592	-8.7
53	627	-7.1
54	680	-7.1
55	694	-7.8
56	704	-7.1
57	712	-7.1
58	749	-8.6
59	750	-7.8
60	773	-7.2
61	777	-7.1
62	796	-7.7
63	797	-8
64	798	-7
65	816	-7.7
66	817	-7.8
67	818	-7.6
68	821	-7.5
69	825	-7.3
70	826	-7.9
71	827	-7
72	829	-7.5
73	906	-7
74	933	-7.7
75	940	-7.2
76	941	-7.4
77	961	-7.1
78	967	-7.1
79	993	-7.5
80	996	-7.1
81	1004	-7.2
82	1019	-7.5
83	1032	-7.3
84	1042	-7.5
85	1054	-7.4
86	1061	-7
87	1065	-7.7
88	1074	-8.1
89	1077	-7.1
90	1083	-8
91	1084	-8
92	1085	-7.4
93	1091	-7.6
94	1096	-7.2
95	1104	-7.3
96	1105	-7
97	1107	-7.1
98	1112	-7.5
99	1113	-7.3
100	1116	-7.4

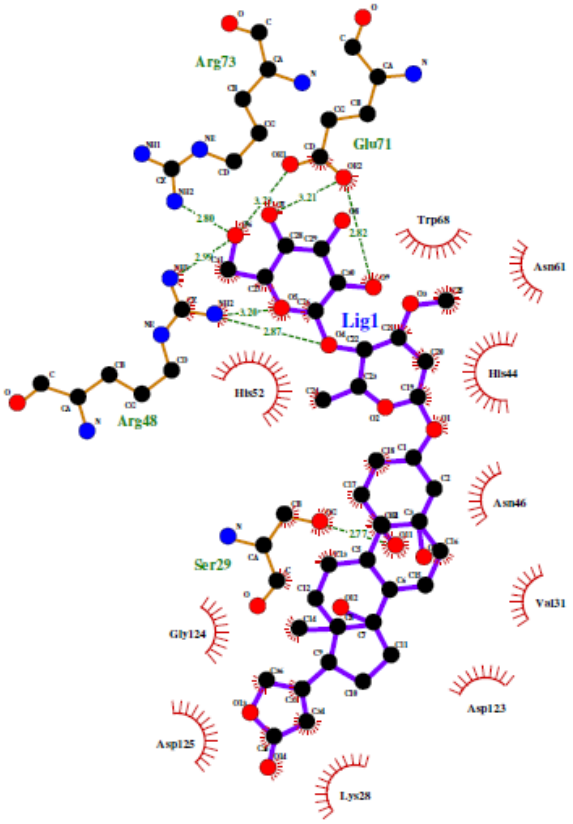
SL No.	NCI_Hits	Best affinity (Kcal/mol)
101	1123	-7.3
102	1145	-7
103	1148	-7.5
104	1153	-7.3
105	1174	-7.8
106	1183	-7.5
107	1188	-8.1
108	1199	-7.3
109	1210	-7.1
110	1228	-7.4
111	1244	-8.1
112	1245	-7.1
113	1246	-7.1
114	1247	-7
115	1259	-7.8
116	1260	-8.6
117	1261	-8.3
118	1272	-7.7
119	1284	-7
120	1286	-7.4
121	1317	-7.5
122	1329	-7.2
123	1375	-7.2
124	1377	-7
125	1396	-7
126	1438	-7.1
127	1449	-7
128	1450	-7.1
129	1453	-7.7
130	1466	-7.7
131	1490	-7.4
132	1529	-7.4
133	1533	-8.1
134	1536	-7.9
135	1543	-7.1
136	1553	-7.2
137	1570	-7.1
138	1571	-7
139	1582	-7.1
140	1586	-8
141	1587	-7.8
142	1588	-7.5
143	1590	-8.3
144	1591	-7.6
145	1604	-7
146	1607	-7
147	1617	-7.4
148	1619	-7.8
149	1644	-7.4
150	1645	-7.5

SL No.	NCI_Hits	Best affinity (Kcal/mol)
151	1646	-7.4
152	1647	-7.2
153	1648	-8.4
154	1649	-8.1
155	1650	-7.9
156	1651	-7.4
157	1652	-7.3
158	1653	-7.1
159	1654	-7.3
160	1655	-7.5
161	1656	-7.3
162	1657	-7.4
163	1658	-7.2
164	1659	-7
165	1672	-7
166	1675	-7.3
167	1676	-7.5
168	1687	-7.4
169	1689	-7.3
170	1694	-7.4
171	1701	-7.7
172	1702	-7.7
173	1705	-7.1
174	1706	-7.2
175	1707	-7.9
176	1709	-7.5
177	1710	-7.9
178	1713	-8.2
179	1718	-7
180	1719	-7.6
181	1720	-7
182	1721	-7
183	1730	-7.4
184	1731	-7.1
185	1733	-7.6
186	1734	-7.2
187	1735	-7.3
188	1746	-7.9
189	1747	-7.1
190	1753	-7.1
191	1754	-7.4
192	1755	-7.5
193	1756	-7.6
194	1759	-8.1
195	1760	-7.9
196	1762	-7.6
197	1768	-7
198	1769	-7.5
199	1772	-7
200	1773	-7.2

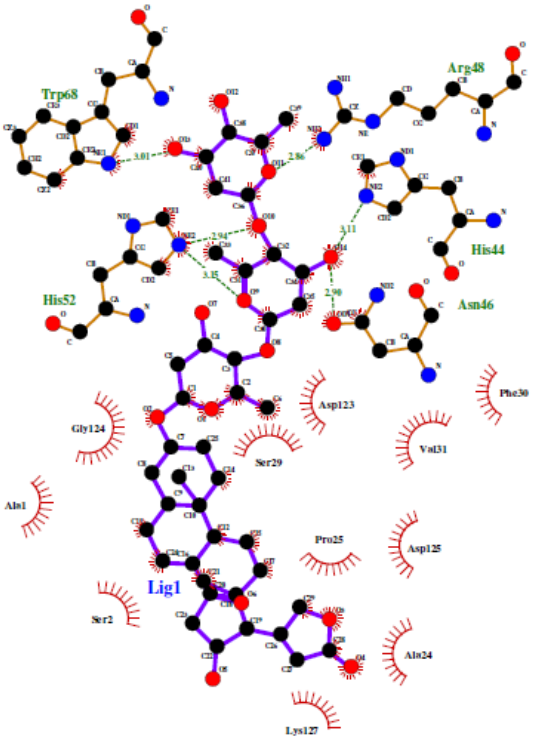
SL No.	NCI_Hits	Best affinity (Kcal/mol)
201	1782	-7.4
202	1789	-7.7
203	1791	-7
204	1792	-7
205	1793	-7.2
206	1799	-7
207	1820	-7.3
208	1821	-7.4
209	1824	-7.1
210	1826	-7
211	1850	-7.5
212	1854	-7.3
213	1855	-7
214	1863	-7.8
215	1869	-8
216	1881	-7.2
217	1884	-7
218	1887	-7.6
219	1890	-7.1
220	1891	-7.1
221	1901	-7.1
222	1903	-7.7
223	1904	-7.6
224	1905	-7
225	1906	-7.6
226	1909	-7.7
227	1911	-7
228	1921	-7.1
229	1922	-7.1
230	1923	-7.1
231	1924	-7.1
232	1925	-7
233	1926	-7.8
234	1927	-7.6
235	1928	-7.4
236	1929	-7.3
237	1930	-7.5
238	1936	-7.2
239	1937	-7.4
240	1938	-9
241	1958	-7.6
242	1963	-7
243	1970	-7.8
244	1973	-7.3
245	1974	-7.4
246	1981	-8.1
247	1992	-8.3
248	2004	-7.3
249	2005	-7.2
250	2019	-7.8

SL No.	NCI_Hits	Best affinity (Kcal/mol)
251	2026	-7.1
252	2028	-7.1
253	2030	-7.2
254	2036	-7
255	2038	-7.1
256	2042	-7.8
257	2045	-7.2
258	2059	-7
259	2062	-7.6
260	2067	-7
261	2073	-7.1
262	2078	-7.7
263	2081	-7
264	2083	-7.6
265	2086	-7.2
266	2088	-8.5
267	2092	-7.7
268	2093	-7.1
269	2094	-8.2
270	2095	-7.6
271	2100	-7.5
272	2102	-7
273	2104	-7.9
274	2105	-7.8
275	2106	-7.6
276	2108	-7.1
277	2110	-7.5
278	2111	-7
279	2114	-7.4
280	2116	-7.1
281	2118	-7.4
282	2119	-7.7
283	2128	-7
284	2131	-7.3
285	2132	-7.2
286	2136	-7.6
287	2137	-7.5
288	2138	-7
289	2140	-7.1
290	2141	-7.6
291	2142	-7.6
292	2145	-7.2
293	2147	-8.1
294	2149	-7.7
295	2150	-7.7
296	2152	-7.8
297	2153	-8.1
298	2154	-8.2
299	2163	-7.2
300	2164	-7.3

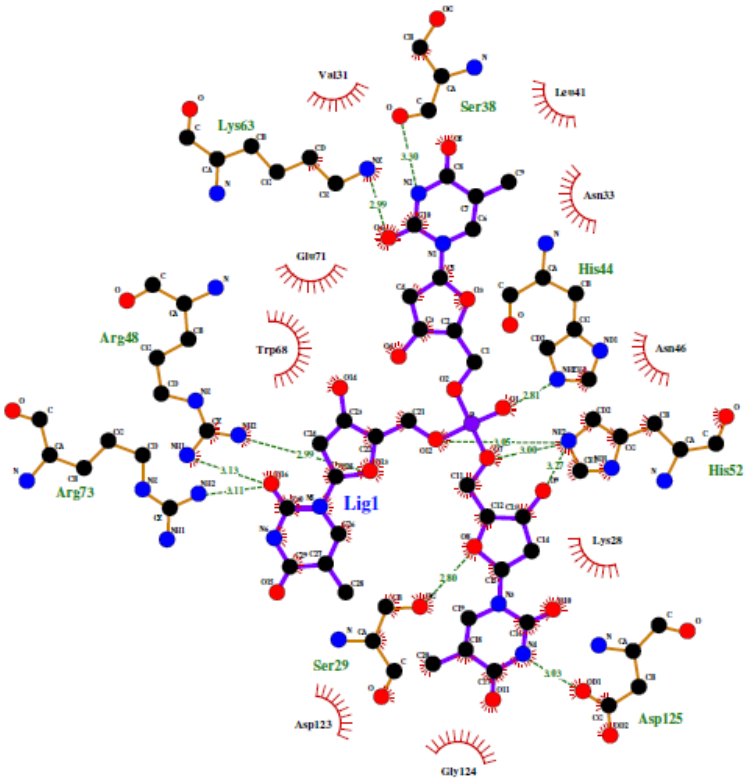
SL No.	NCI_Hits	Best affinity (Kcal/mol)
301	2165	-7.4
302	2166	-7
303	2175	-7.7
304	2177	-7.1
305	2178	-7.5
306	2179	-7.6
307	2180	-7.7
308	2181	-7.3
309	2188	-7.1
310	2189	-7.1
311	2196	-7.9
312	2204	-7.2
313	2207	-7.2
314	2209	-7.5
315	2215	-7.4
316	2217	-7.5



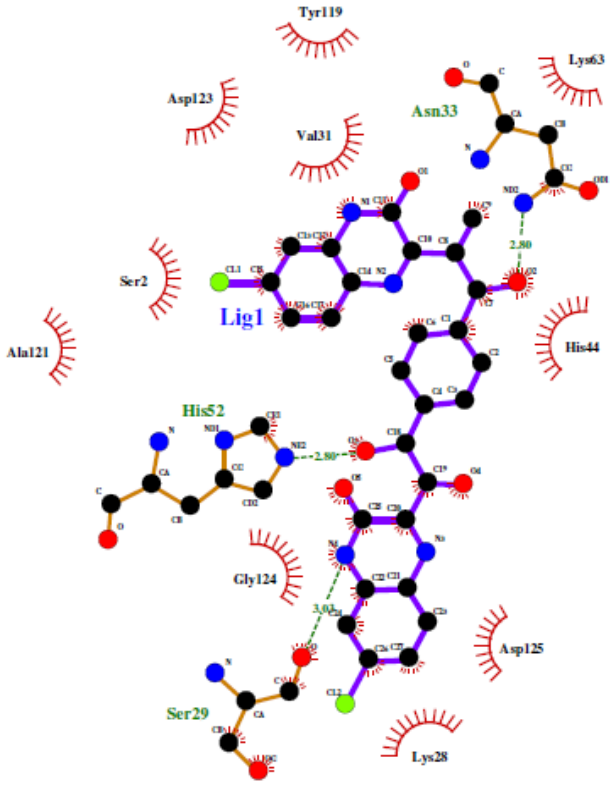
GAL1_NCI_50



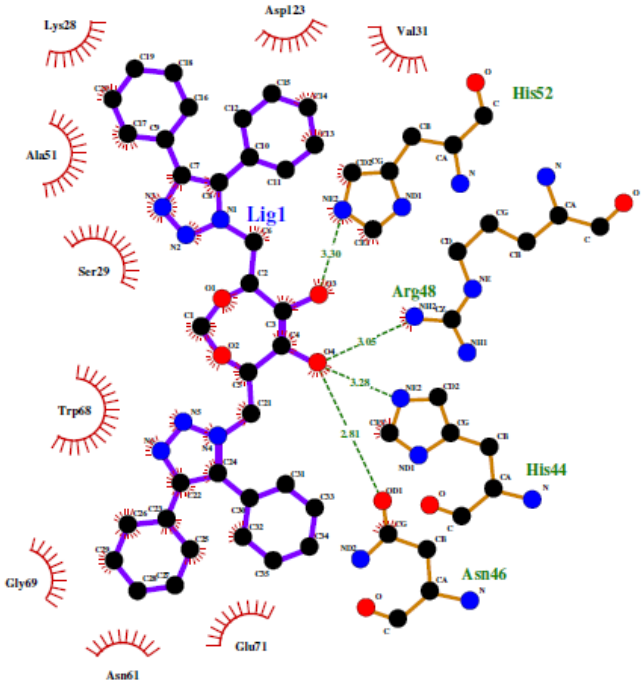
GAL1_NCI_379



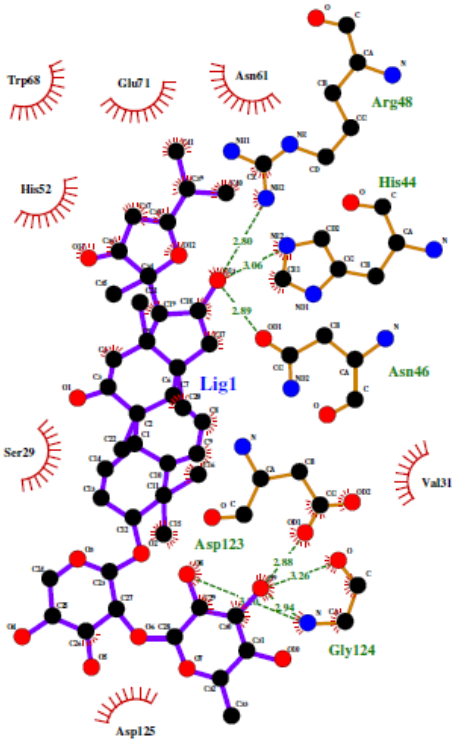
GAL1_NCI_592



GAL1_NCI_1713



GAL1_NCI_1938



GAL1_NCI_1992

Figure 9. Galectin-1 and ligand interaction analysis using LigPlot+ v1.4.5

4.7 QSAR STUDIES:

Quantitative Structure-Activity Relationship (QSAR) is used to build mathematical models for bioactivity prediction. Any QSAR model for given set of molecules correlates the activities with properties inherent to each molecule in the set itself. 35 compounds of CHEMBL database with known IC_{50} values were used to generate independent training and test data sets. Initially, all 35 compounds were clustered and defined as training sets, using LigandScout software. 7 cluster were formed and 28 considered as training set (Figure 10). Training set and test set descriptors, generated by PaDEL Descriptor software were used to construct and cross validate the QSAR model. QSAR model equation with best score (q2) and correlation (r^2) value of the fits were reported (Table 4). The QSAR model equation was validated with experimental IC50 values of training set (Table 5) (Figure 11) and then, used for the prediction of activities of test set compounds (Table 6).

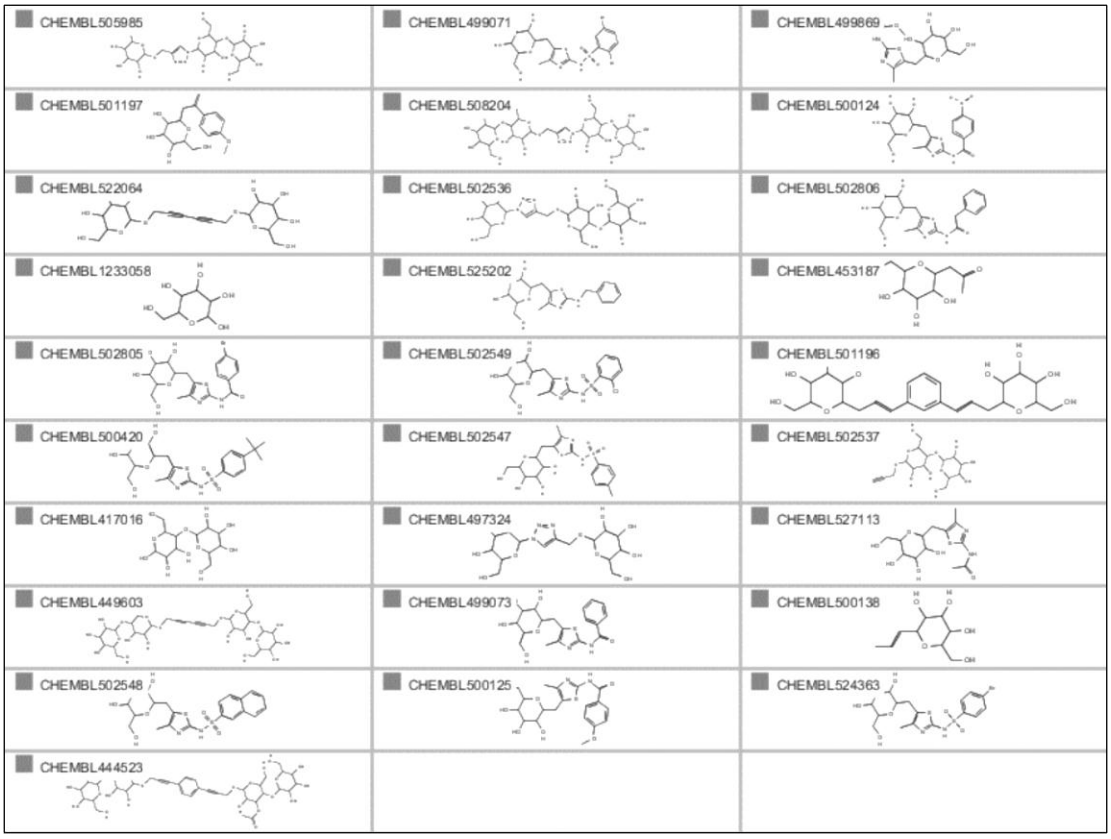


Figure 10. Training set compounds for QSAR model generation & validation

Table 4. QSAR equation generation

QSAR Model	Best correlation (r^2)	Best score (q^2)
pIC_{50} = mul(plus(const(0.000661023), neg(plus(plus(sqrt(neg(spline(7.79542,nssCH2))), mul(plus(const(-528488),mul(const(-0.0164224), neg(neg(qspline(-5750.98,LipoaffinityIndex))))), neg(qspline(1.28571e+006,LipoaffinityIndex)))), gauss(0.11092,-0.547383,0.00676317,RotBFrac)))), neg(plus(const(-1.19405e+006),mul(const(-0.119312), neg(plus(const(-3.35299e+006),mul(const(-0.0221126), neg(qspline(-24658.6,LipoaffinityIndex))))))))	0.850138	0.86952

mul= multiplication, plus= addition, const= constant, neg= negation, sqrt= squareroot, spline(knot,descriptor_name)= returns the value substracted by knot value, nssCH2= count of atom-type E-state: -CH2-, qspline= quadraticspline, lipoaffinityindex= lipoaffinity index, gauss= Gaussian function in 1D, RotBFrac= rotatable bond count.

Table 5. QSAR Model Validation

compounds	Predicted IC50 (nM)	Experimental IC50 (nM)
CHEMBL505985	587570.5	300000
CHEMBL499071	4.18E+06	5.00E+06
CHEMBL499869	3.23E+06	5.00E+06
CHEMBL501197	6.96E+06	5.00E+06
CHEMBL508204	232830.6	300000
CHEMBL500124	3.68E+06	5.00E+06
CHEMBL522064	1.77E+06	5.00E+06
CHEMBL502536	585544.6	600000
CHEMBL502806	5.39E+06	5.00E+06
CHEMBL1233058	4.94E+07	5.00E+07
CHEMBL525202	6.61E+06	5.00E+06
CHEMBL453187	3.13E+06	5.00E+06
CHEMBL502805	5.07E+06	2.50E+06

CHEMBL502549	4.36E+06	5.00E+06
CHEMBL501196	4.33E+06	5.00E+06
CHEMBL500420	7.10E+06	5.00E+06
CHEMBL502547	4.87E+06	5.00E+06
CHEMBL502537	1.48E+06	600000
CHEMBL417016	1.14E+06	800000
CHEMBL497324	1.42E+06	5.00E+06
CHEMBL527113	3.37E+06	5.00E+06
CHEMBL449603	297824.2	5.00E+06
CHEMBL499073	4.98E+06	5.00E+06
CHEMBL500138	4.95E+06	5.00E+06
CHEMBL502548	6.32E+06	5.00E+06
CHEMBL500125	5.05E+06	5.00E+06
CHEMBL524363	4.40E+06	5.00E+06
CHEMBL444523	571689.2	300000

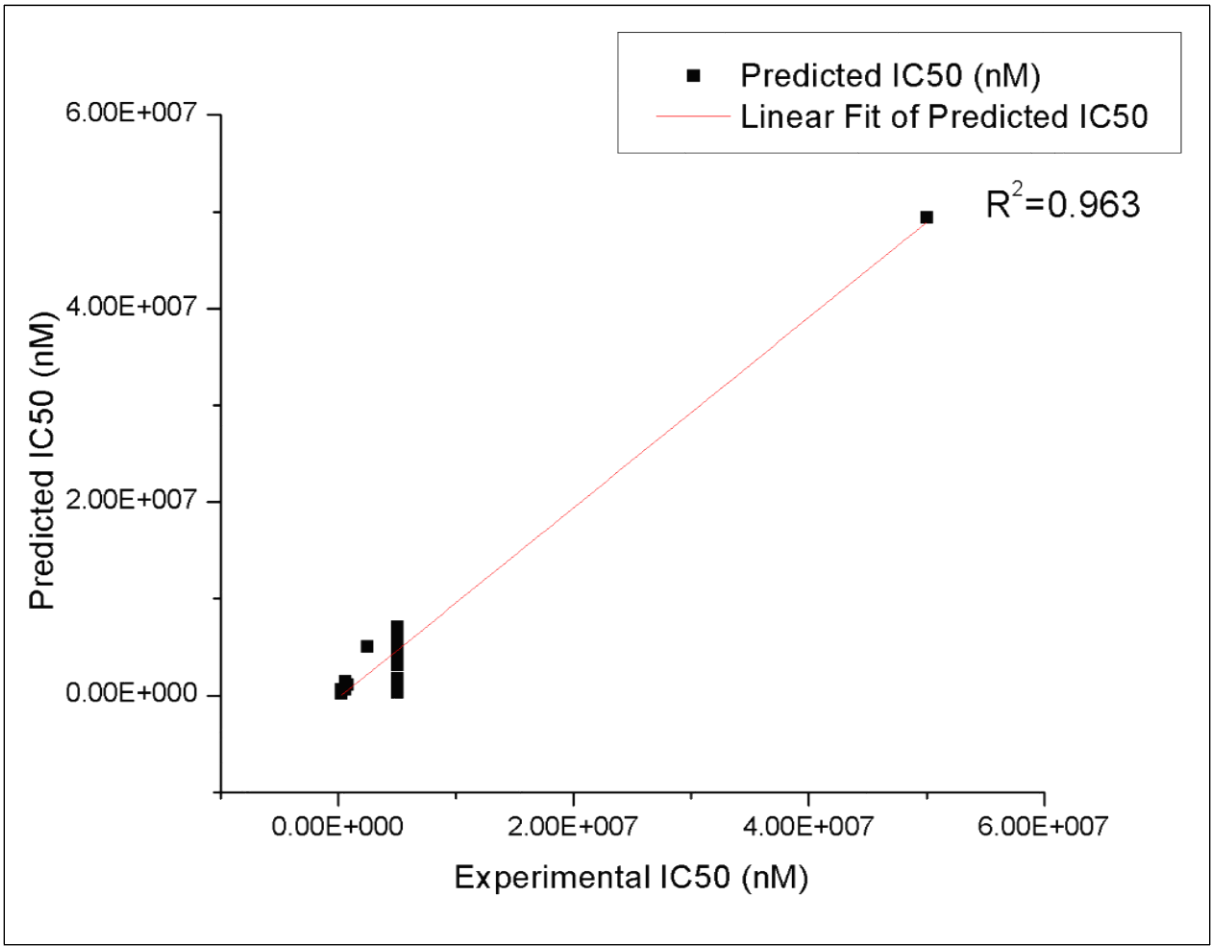
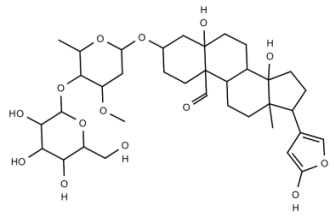
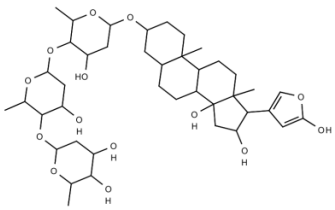
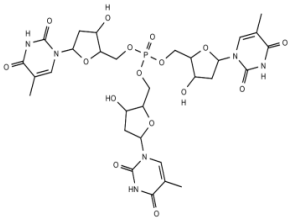
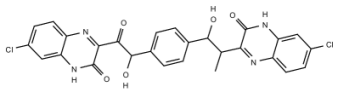
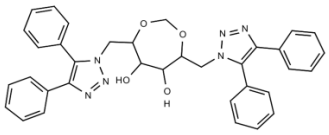
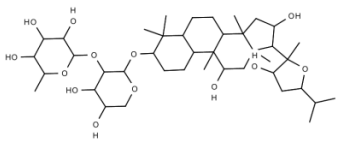


Figure 11. QSAR model validation plot between predicted and experimental IC50 values

Table 6. QSAR Prediction & Selection of Molecules:

Sl no.	NCI hit	STRUCTURE	NSC	Binding energy (Kcal/mol)	clusters (max20)	Predicted IC50(mM)
1	50		4320	-8.0	18	8.537486
2	379		95099	-8.1	14	7.612021
3	592		140924	-8.7	20	18.053857
4	1713		641818	-8.2	17	8.623231
5	1938		683367	-9.0	13	8.033781
6	1992		694346	-8.3	18	8.622238

CHAPTER 5

CONCLUSION

CONCLUSIONS:

Based on our results, we obtained 6 potential inhibitors of galectin-1 from NCI compound database. These molecules are named NSC4320, NSC95099, NSC140924, NSC641818, NSC683367 and NSC694346 having good binding energy (≥ -8.0 Kcal/mol) with better cluster (≥ 13). Our inhibitors preferably interacted with galectin-1 carbohydrate recognition domain residues that play major roles in adherence between HIV-1 capsid protein (gp120) and CD4+ helper T cells. We also predicted inhibitory concentration (IC50) of inhibitors which are better than natural inhibitor (lactose) e.g. less than 10mM except NSC140924

REFERENCES:

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY: **Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010**. *The Lancet* 2013, **380**(9859):2095-2128.
2. Turner BG, Summers MF: **Structural biology of HIV**. *Journal of molecular biology* 1999, **285**(1):1-32.
3. Capon DJ, Ward RH: **The CD4-gpl20 interaction and AIDS pathogenesis**. *Annual review of immunology* 1991, **9**(1):649-678.
4. Eiden LE, Lifson JD: **HIV interactions with CD4: a continuum of conformations and consequences**. *Immunology today* 1992, **13**(6):201-206.
5. Bour S, Geleziunas R, Wainberg MA: **The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection**. *Microbiological reviews* 1995, **59**(1):63-93.
6. Chan DC, Kim PS: **HIV entry and its inhibition**. *Cell* 1998, **93**(5):681-684.
7. Gallo SA, Finnegan CM, Viard M, Raviv Y, Dimitrov A, Rawat SS, Puri A, Durell S, Blumenthal R: **The HIV Env-mediated fusion reaction**. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 2003, **1614**(1):36-50.
8. Feinberg H, Mitchell DA, Drickamer K, Weis WI: **Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR**. *Science* 2001, **294**(5549):2163-2166.
9. Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J, Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR: **DC-SIGN, a Dendritic Cell–Specific HIV-1-Binding Protein that Enhances trans-infection of T Cells**. *cell* 2000, **100**(5):587-597.
10. Kwon DS, Gregorio G, Bitton N, Hendrickson WA, Littman DR: **DC-SIGN-Mediated Internalization of HIV Is Required for Trans-enhancement of T Cell Infection**. *Immunity* 2002, **16**(1):135-144.
11. Lambert AA, Gilbert C, Richard M, Beaulieu AD, Tremblay MJ: **The C-type lectin surface receptor DCIR acts as a new attachment factor for HIV-1 in dendritic cells and contributes to trans-and cis-infection pathways**. *Blood* 2008, **112**(4):1299-1307.
12. Lin G, Simmons G, Pöhlmann S, Baribaud F, Ni H, Leslie GJ, Haggarty BS, Bates P, Weissman D, Hoxie JA: **Differential N-linked glycosylation of human immunodeficiency virus and Ebola virus envelope glycoproteins modulates interactions with DC-SIGN and DC-SIGNR**. *Journal of virology* 2003, **77**(2):1337-1346.
13. McDonald D, Wu L, Bohks SM, KewalRamani VN, Unutmaz D, Hope TJ: **Recruitment of HIV and its receptors to dendritic cell-T cell junctions**. *Science* 2003, **300**(5623):1295-1297.
14. Mercier S, St-Pierre C, Pelletier I, Ouellet M, Tremblay MJ, Sato S: **Galectin-1 promotes HIV-1 infectivity in macrophages through stabilization of viral adsorption**. *Virology* 2008, **371**(1):121-129.
15. Ouellet M, Mercier S, Pelletier I, Bounou S, Roy J, Hirabayashi J, Sato S, Tremblay MJ: **Galectin-1 acts as a soluble host factor that promotes HIV-1 infectivity through stabilization of virus attachment to host cells**. *The Journal of Immunology* 2005, **174**(7):4120-4126.

16. St-Pierre C, Manya H, Ouellet M, Clark GF, Endo T, Tremblay MJ, Sato S: **Host-soluble galectin-1 promotes HIV-1 replication through a direct interaction with glycans of viral gp120 and host CD4**. *Journal of virology* 2011, **85**(22):11742-11751.

17. St-Pierre C, Ouellet M, Tremblay MJ, Sato S: **Chapter Thirteen-Galectin-1 and HIV-1 Infection**. *Methods in enzymology* 2010, **480**:267-294.

18. Turville S, Wilkinson J, Cameron P, Dable J, Cunningham AL: **The role of dendritic cell C-type lectin receptors in HIV pathogenesis**. *Journal of leukocyte biology* 2003, **74**(5):710-718.

19. Nio-Kobayashi J, Takahashi-Iwanaga H, Iwanaga T: **Immunohistochemical localization of six galectin subtypes in the mouse digestive tract**. *Journal of Histochemistry & Cytochemistry* 2009, **57**(1):41-50.

20. Rabinovich G, Castagna L, Landa C, Riera CM, Sotomayor C: **Regulated expression of a 16-kd galectin-like protein in activated rat macrophages**. *Journal of leukocyte biology* 1996, **59**(3):363-370.

21. Stillman BN, Hsu DK, Pang M, Brewer CF, Johnson P, Liu F-T, Baum LG: **Galectin-3 and galectin-1 bind distinct cell surface glycoprotein receptors to induce T cell death**. *The Journal of Immunology* 2006, **176**(2):778-789.

22. Barondes SH, Cooper DN, Gitt MA, Leffler H: **Galectins. Structure and function of a large family of animal lectins**. *Journal of Biological Chemistry* 1994, **269**:20807-20807.

23. Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Muller WE: **Oligosaccharide specificity of galectins: a search by frontal affinity chromatography**. *Biochimica et Biophysica Acta (BBA)-General Subjects* 2002, **1572**(2):232-254.

24. Hirabayashi J, Kasai K-i: **The family of metazoan metal-independent β-galactoside-binding lectins: structure, function and molecular evolution**. *Glycobiology* 1993, **3**(4):297-304.

25. Leffler H, Carlsson S, Hedlund M, Qian Y, Poirier F: **Introduction to galectins**. *Glycoconjugate journal* 2002, **19**(7-9):433-440.

26. Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH: **The origins of acquired immune deficiency syndrome viruses: where and when?** *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2001, **356**(1410):867-876.

27. Bailes E, Chaudhuri RR, Santiago ML, Bibollet-Ruche F, Hahn BH, Sharp PM: **The evolution of primate lentiviruses and the origins of AIDS**. In: *The molecular epidemiology of human viruses.* Springer; 2002: 65-96.

28. Freed EO: **HIV-1 replication**. *Somatic cell and molecular genetics* 2001, **26**(1-6):13-33.

29. Ahmad N, Gabius H-J, Sabesan S, Oscarson S, Brewer CF: **Thermodynamic binding studies of bivalent oligosaccharides to galectin-1, galectin-3, and the carbohydrate recognition domain of galectin-3**. *Glycobiology* 2004, **14**(9):817-825.

30. Danguy A, Camby I, Kiss R: **Galectins and cancer**. *Biochimica et Biophysica Acta (BBA)-General Subjects* 2002, **1572**(2):285-293.

31. Liu F-T, Rabinovich GA: **Galectins as modulators of tumour progression**. *Nature Reviews Cancer* 2005, **5**(1):29-41.

32. Hughes RC: **Secretion of the galectin family of mammalian carbohydrate-binding proteins**. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1999, **1473**(1):172-185.
33. López-Lucendo MF, Solís D, André S, Hirabayashi J, Kasai K-i, Kaltner H, Gabius H-J, Romero A: **Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding**. *Journal of molecular biology* 2004, **343**(4):957-970.
34. Liao D-I, Kapadia G, Ahmed H, Vasta GR, Herzberg O: **Structure of S-lectin, a developmentally regulated vertebrate beta-galactoside-binding protein**. *Proceedings of the National Academy of Sciences* 1994, **91**(4):1428-1432.
35. Ford MG, Weimar T, Köhli T, Woods RJ: **Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity**. *Proteins: Structure, Function, and Bioinformatics* 2003, **53**(2):229-240.
36. Miller M, Nesmelova I, Platt D, Klyosov A, Mayo K: **The carbohydrate-binding domain on galectin-1 is more extensive for a complex glycan than for simple saccharides: implications for galectin-glycan interactions at the cell surface**. *Biochem J* 2009, **421**:211-221.
37. Baum LG, Pang M, Perillo NL, Wu T, Delegeane A, Uittenbogaart CH, Fukuda M, Seilhamer JJ: **Human thymic epithelial cells express an endogenous lectin, galectin-1, which binds to core 2 O-glycans on thymocytes and T lymphoblastoid cells**. *The Journal of experimental medicine* 1995, **181**(3):877-887.
38. Zuñiga E, Rabinovich GA, Iglesias MM, Gruppi A: **Regulated expression of galectin-1 during B-cell activation and implications for T-cell apoptosis**. *Journal of leukocyte biology* 2001, **70**(1):73-79.
39. Blaser C, Kaufmann M, Müller C, Zimmermann C, Wells V, Mallucci L, Pircher H: **β-Galactoside-binding protein secreted by activated T cells inhibits antigen-induced proliferation of T cells**. *European journal of immunology* 1998, **28**(8):2311-2319.
40. Dettin L, Rubinstein N, Aoki A, Rabinovich GA, Maldonado CA: **Regulated expression and ultrastructural localization of galectin-1, a proapoptotic β-galactoside-binding lectin, during spermatogenesis in rat testis**. *Biology of reproduction* 2003, **68**(1):51-59.
41. Jeschke U, Reimer T, Bergemann C, Wiest I, Schulze S, Friese K, Walzel H: **Binding of galectin-1 (gal-1) on trophoblast cells and inhibition of hormone production of trophoblast tumor cells in vitro by gal-1**. *Histochemistry and cell biology* 2004, **121**(6):501-508.
42. Toscano MA, Campagna L, Molinero LL, Cerliani JP, Croci DO, Ilarregui JM, Fuertes MB, Nojek IM, Fededa JP, Zwirner NW: **Nuclear factor (NF)-κB controls expression of the immunoregulatory glycan-binding protein galectin-1**. *Molecular immunology* 2011, **48**(15):1940-1949.
43. Singh BN, Hayes GR, Lucas JJ, Sommer U, Viseux N, Mirgorodskaya E, Trifonova RT, Sassi RRS, Costello CE, Fichorova RN: **Structural details and composition of Trichomonas vaginalis lipophosphoglycan in relevance to the epithelial immune function**. *Glycoconjugate journal* 2009, **26**(1):3-17.
44. Fichorova RN: **Impact of< i> T. vaginalis</i> infection on innate immune responses and reproductive outcome**. *Journal of reproductive immunology* 2009, **83**(1):185-189.
45. Imbeault M, Lodge R, Ouellet M, Tremblay MJ: **Efficient magnetic bead-based separation of HIV-1-infected cells using an improved reporter virus system**

reveals that p53 up-regulation occurs exclusively in the virus-expressing cell population. *Virology* 2009, **393**(1):160-167.
46. de Witte L, Nabatov A, Pion M, Fluitsma D, De Jong MA, de Gruijl T, Piguet V, van Kooyk Y, Geijtenbeek TB: **Langerin is a natural barrier to HIV-1 transmission by Langerhans cells**. *Nature medicine* 2007, **13**(3):367-371.
47. Piguet V, Steinman RM: **The interaction of HIV with dendritic cells: outcomes and pathways**. *Trends in immunology* 2007, **28**(11):503-510.
48. Wu L, KewalRamani VN: **Dendritic-cell interactions with HIV: infection and viral dissemination**. *Nature Reviews Immunology* 2006, **6**(11):859-868.
49. Bates EE, Fournier N, Garcia E, Valladeau J, Durand I, Pin J-J, Zurawski SM, Patel S, Abrams JS, Lebecque S: **APCs express DCIR, a novel C-type lectin surface receptor containing an immunoreceptor tyrosine-based inhibitory motif**. *The Journal of Immunology* 1999, **163**(4):1973-1983.
50. Tremblay MJ, Fortin J-F, Cantin R: **The acquisition of host-encoded proteins by nascent HIV-1**. *Immunology today* 1998, **19**(8):346-351.
51. Cantin R, Méthot S, Tremblay MJ: **Plunder and stowaways: incorporation of cellular proteins by enveloped viruses**. *Journal of virology* 2005, **79**(11):6577-6587.
52. Hildreth J, Orentas RJ: **Involvement of a leukocyte adhesion receptor (LFA-1) in HIV-induced syncytium formation**. *Science* 1989, **244**(4908):1075-1078.
53. Bobardt MD, Saphire A, Hung H-C, Yu X, Van der Schueren B, Zhang Z, David G, Gallay PA: **Syndecan captures, protects, and transmits HIV to T lymphocytes**. *Immunity* 2003, **18**(1):27-39.
54. Gallay P: **Syndecans and HIV-1 pathogenesis**. *Microbes and infection* 2004, **6**(6):617-622.
55. de Witte L, Bobardt M, Chatterji U, Degeest G, David G, Geijtenbeek TB, Gallay P: **Syndecan-3 is a dendritic cell-specific attachment receptor for HIV-1**. *Proceedings of the National Academy of Sciences* 2007, **104**(49):19464-19469.
56. St-Pierre C, Ouellet M, Giguère D, Ohtake R, Roy R, Sato S, Tremblay MJ: **Galectin-1-specific inhibitors as a new class of compounds to treat HIV-1 infection**. *Antimicrobial agents and chemotherapy* 2012, **56**(1):154-162.
57. Sato S, Ouellet M, St-Pierre C, Tremblay MJ: **Glycans, galectins, and HIV-1 infection**. *Annals of the New York Academy of Sciences* 2012, **1253**(1):133-148.
58. Liao C, Sitzmann M, Pugliese A, Nicklaus MC: **Software and resources for computational medicinal chemistry**. *Future medicinal chemistry* 2011, **3**(8):1057-1085.
59. Oprea TI, Matter H: **Integrating virtual screening in lead discovery**. *Current opinion in chemical biology* 2004, **8**(4):349-358.
60. Congreve M, Murray CW, Blundell TL: **Keynote review: Structural biology and drug discovery**. *Drug Discovery Today* 2005, **10**(13):895-907.
61. Jorgensen WL: **The many roles of computation in drug discovery**. *Science* 2004, **303**(5665):1813-1818.
62. Kubinyi H: **Success stories of computer-aided design**. *Computer Applications in Pharmaceutical Research and Development* 2006:377-424.
63. Alvarez J, Shoichet B: **Virtual screening in drug discovery**: CRC press; 2005.
64. Hou T, Xu X: **Recent development and application of virtual screening in drug discovery: an overview**. *Current pharmaceutical design* 2004, **10**(9):1011-1033.
65. Keri G, Orfi L, Eros D, Hegymegi-Barakonyi B, Szantai-Kis C, Horvath Z, Waczek F, Marosfalvi J, Szabadkai I, Pato J: **Signal transduction therapy with rationally designed kinase inhibitors**. *Current Signal Transduction Therapy* 2006, **1**(1):67-95.

66. Shoichet BK: **Virtual screening of chemical libraries**. *Nature* 2004, **432**(7019):862-865.
67. Lipinski C, Hopkins A: **Navigating chemical space for biology and medicine**. *Nature* 2004, **432**(7019):855-861.
68. Lu Y, Nikolovska-Coleska Z, Fang X, Gao W, Shangary S, Qiu S, Qin D, Wang S: **Discovery of a nanomolar inhibitor of the human murine double minute 2 (MDM2)-p53 interaction through an integrated, virtual database screening strategy**. *Journal of medicinal chemistry* 2006, **49**(13):3759-3762.
69. Lyne PD, Kenny PW, Cosgrove DA, Deng C, Zabludoff S, Wendoloski JJ, Ashwell S: **Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening**. *Journal of medicinal chemistry* 2004, **47**(8):1962-1968.
70. Langer T, Krovat E: **Chemical feature-based pharmacophores and virtual library screening for discovery of new leads**. *Current opinion in drug discovery & development* 2003, **6**(3):370-376.
71. Lloyd DG, Golfis G, Knox AJ, Fayne D, Meegan MJ, Oprea TI: **Oncology exploration: charting cancer medicinal chemistry space**. *Drug discovery today* 2006,**11**(3):149-159.
72. **NCI Open Database Compounds, release 4**; National Cancer Institute, NationalInstitutes of Health: Marc C. Nicklaus - Head, CADD Group, CBL, Center for Cancer Research, NCI, NIH). Available online at: http://cactus.nci.nih.gov/download/nci (accessed May.2012).
73. Wolber G, Langer T: **LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters**. *Journal of chemical information and modeling* 2005, **45**(1):160-169.
74. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov IN, Bourne PE: **The protein data bank**. *Nucleic acids research* 2000, **28**(1):235-242.
75. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera—a visualization system for exploratory research and analysis**. *Journal of computational chemistry* 2004, **25**(13):1605-1612.
76. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ: **GROMACS: fast, flexible, and free**. *Journal of computational chemistry* 2005, **26**(16):1701-1718.
77. Trott O, Olson AJ: **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading**. *Journal of computational chemistry* 2010, **31**(2):455-461.
78. Laskowski RA, Swindells MB: **LigPlot+: multiple ligand–protein interaction diagrams for drug discovery**. *Journal of chemical information and modeling* 2011, **51**(10):2778-2786.
79. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B: **ChEMBL: a large-scale bioactivity database for drug discovery**. *Nucleic acids research* 2012, **40**(D1):D1100-D1107.
80. Vainio MJ, Johnson MS: **McQSAR: a multiconformational quantitative structure-activity relationship engine driven by genetic algorithms**. *Journal of chemical information and modeling* 2005, **45**(6):1953-1961.
81. Yap CW: **PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints**. *Journal of computational chemistry* 2011, **32**(7):1466-1474.